Progressive Ratio and Delay Discounting Assessments for Young Children
With and Without Problem Behavior

By

Emily S. Weaver

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Special Education

August, 9, 2019

Nashville, Tennessee

Approved:

Blair P. Lloyd, Ph.D. BCBA-D

Joseph H. Wehby, Ph.D.

Paul J. Yoder, Ph.D.

Julie Lounds Taylor, Ph.D.

Johanna L. Staubitz, Ph.D.

To my parents, Mary and Tom Weaver, my first and best teachers

and

To Will Madden, who reminds me to look at ducks from time to time

TABLE OF CONTENTS

Page

DEDICATION ii

LIST OF TABLES v

LIST OF FIGURES vi

Chapter

1. Introduction 1

   Progressive Ratio Assessments 3
   Delay Discounting Assessments 6
   Rationale 11
   Purpose 12

2. Method 14

   Inclusion and Exclusion Criteria 14
   Recruitment 15
   Participants 16
   Setting 17
   Materials 17
   Assessment Schedule 18
   Procedures 19
   Research Assistant Training 26
   Measures and Outcomes 27
   Inter-Observer Agreement 28
   Procedural Fidelity 30
   Data Analysis 32

3. Results 36

   Research Question 1: Dependability of Progressive Ratio Scores 38
   Research Question 2: Distribution and Correlations for Delay Discounting Scores 40
   Research Question 3: Associations Between Breakpoint and Problem Behavior 44
   Research Question 4: Associations Between Delay Discounting and Problem Behavior 44
   Research Question 5: Correlations Between Breakpoints and Delay Discounting 46

4. Discussion.. 47

Progressive Ratio.. 47
Delay Discounting.. 49
Limitations.. 51
Future Directions.. 53

REFERENCES.. 56

Appendix

A. Example Math Worksheet for Progressive Ratio Academic Task 61

B. Example Data Collection Sheet ADT-5.. 62

C. ADT-5 Script 63

D. Introductory Scripts for Progressive Ratio Assessments.. 64

LIST OF TABLES

Table Page

1. Inclusion and Exclusion Criteria and Information Source by Group 15
2. Characteristics of Included Participants by Subgroup and Assessment Type 16
3. Selected Demographics of School Settings 17
4. Trials, Delays and Outcomes for Delay Discounting Assessments 25
5. Progressive Ratio Assessment Reliability 29
6. Progressive Ratio Procedural Fidelity 31
7. Total and Subscale SAEBRS Scores by Participant Groups 36
8. Pearson's *r* Correlations Among Study Variables 39

LIST OF FIGURES

Figure Page

1. Response-Guided Changes to Delays and Outcomes for the ADT-5 23
2. Comparison of Means and Standard Deviations of Breakpoints 37
3. Comparison of Means and Standard Deviations of log(*k*) Values 37
4. Demonstration of the Change in Absolute *g* When Tasks Are Added 40
5. Distribution of *k* Values Obtained from the ADT-5 41
6. Box Plots of log(*k*) Values Obtained from the ADT-5 42
7. Associations between *k* values from the ADT-5 with Different Rewards 43
8. Graphs of Consumption as a Function of Schedule in Progressive Ratio Assessments 45

CHAPTER 1

INTRODUCTION

For decades, reinforcement has been recognized as an effective component of behavioral interventions in educational settings (e.g., Hall, Lund, & Jackson, 1968; Harris, Wolf, & Baer, 1964; O'Leary, Becker, Evanse, & Saudargas, 1969). Reinforcement contingencies are an important component of several effective school-based interventions designed to improve student's behaviors, including group contingencies (see Maggin, Johnson, Chafouleas, Ruberto, & Berggren, 2012 for a review), positive behavior support programs (see Horner, Sugai, & Anderson, [2010] and Solomon, Klein, Hintze, Cressey, & Peller, [2012] for reviews) and token economies (see Maggin, Chafouleas, Goddard, & Johnson, 2011, for a review). For students with disabilities, reinforcement is a recommended practice for changing school behavior (e.g., DuPaul & Stoner, 2014; Odom, Collet-Klingenberg, Rogers, & Hattan, 2010).

In school settings, teachers can arrange contingencies to reinforce appropriate classroom behavior. They might make preferred stimuli or activities available following desired behaviors, or following the absence of challenging behavior. However, other classroom contingencies compete with these pre-arranged contingencies. For example, a teacher might deliver attention contingent on a student's engagement with academic tasks, but her attention might also be available contingent on refusing the task, and peer attention might be available contingent on disruptive behavior. In cases like these, a student might continue to engage in task refusal or disruptive behavior because of these competing contingencies.

When multiple contingencies are in place, the matching law (Hernstein, 1961) predicts that the relative rates of each behavior will be equal to the relative rates of reinforcement obtained for each behavior. The predictive power of the matching law has been demonstrated in controlled experiments with students in schools. For example, Mace, McCurdy, and Quigley (1990) examined the rates of responding on two different math tasks when contingencies were manipulated for an adolescent with low achievement scores in math. When teachers reinforced both multiplication and division work by delivering rewards after two problems were completed on average, the student solved an equal amount of the two types of problems. However, when teachers changed the rates of reinforcement for one type of problem by delivering a reward following each completed problem, he allocated a higher proportion of his responding to the task reinforced on a richer schedule. Overall, the relative rate of responding on each math task was roughly equal to the relative rate of the reinforcers obtained for each task, as predicted by the matching law.

In authentic classroom settings, however, many more than two contingencies might be in place at the same time, and they might vary by more than rate. Reinforcers of different quality might be concurrently available contingent on differentially effortful responses, with reinforcement at different delays and different rates (Billington & DiTommaso, 2003). All of these dimensions of reinforcement (quality, response effort, delay, and rate), can impact student response allocation. This can complicate efforts to design reinforcement contingencies that improve classroom behavior. Further complicating matters, evidence suggests that the influence of these dimensions varies by individual. In studies where dimensions of reinforcement were manipulated in different concurrently available schedules, response allocation was most strongly influenced by rate for some school-aged children, while for others delay or quality was more influential (Neef & Lutz, 2001; Neef, Shade, & Miller, 1994; Romani, Alcorn, Miller, & Clark, 2017).

In some cases, assessments are conducted to inform individualized behavior interventions, to design therapeutic reinforcement contingencies that are competitive with other contingencies in the classroom. Often, these assessments inform selection of the highest quality reinforcer, by determining which stimuli are the most effective reinforcers when all other dimensions are held constant. These assessments include preference assessments (Hagopian, Long, & Rush, 2004; Paramore & Higbee, 2005), reinforcer assessments (Paclawskyj & Vollmer, 1995) and functional analyses (Broussard & Northup, 1997; Wright-Gallo, Higbee, Reagon, & Davey, 2006), all of which have been used to design effective interventions in school settings. Assessments that address the effects of other dimensions of reinforcement on children's classroom behavior are less common.

For students who are more responsive to manipulations of reinforcement rate, delay, or response effort than manipulations of quality, assessments designed to select the highest-quality reinforcers might not lead to the most effective interventions. Assessing the effects of response effort, delay, and rate might improve the design of individualized interventions for these students. Further, understanding these effects might give teachers and school staff flexibility when designing and choosing interventions for all students. In some cases, manipulating access to the highest quality reinforcer is not feasible in classrooms. For example, teachers might not be able to control the delivery of peer attention contingent on disruptive behaviors, or might not be willing to grant access to loud or distracting tangibles during class time, even if these are the highest quality reinforcers. In these cases, the best practical classroom interventions might involve manipulating other dimensions of reinforcement beyond quality.

The effects of response effort and reinforcement rate on behavior can be measured in a progressive ratio assessment (Roane, 2008). The effects of delays on the subjective value of reinforcers can be measured with a delay discounting assessment (Odum, 2011a). Both of these assessments, along with assessments of reinforcer quality, have potential to inform effective classroom contingencies. However, to be useful, these assessments must have adequate reliability and generalizability. That is, results of these assessments must reflect an individual's stable tendency to behave outside of the assessment context. Some evidence suggests that scores from both assessments reflect stable, within-person constructs. However, the reliability and generalizability of these measures have not been evaluated for school-aged children.

Progressive Ratio Assessments

In a progressive ratio schedule, reinforcers are delivered contingent on a fixed number of responses that systematically increases within a session (Roane, 2008). For example, a progressive ratio schedule might begin with a reinforcer delivered following one response (a fixed-ratio of one response to one reinforcer [FR-1]), with the next reinforcer delivered following two responses (FR-2), the next after three responses (FR-3), and so on (e.g., Glover, Roane, Kadey, & Grow, 2008). Progressive ratio schedules can be used to determine the amount of effort a person is willing to put forth to earn a given reinforcer, indexed by both total responses emitted and breakpoint. The breakpoint is defined as the response value of the last schedule completed before the person stops responding, requests to stop, or reaches a pre-defined session duration limit (Roane, 2008). For example, a child might press a button when each button press produces an M&M (FR-1), and might continue pressing the button as the schedule increases to an FR-2, FR-3, and FR-4. If, after completing the FR-4 schedule, the child stopped pressing the button before beginning the next schedule, his breakpoint would be 4 and his total responses would be 10.

In applied research, breakpoints have been used as an index of reinforcer quality within individuals. In these cases, progressive ratio schedules have been presented to individuals in repeated sessions with different reinforcers earned in each session. Reinforcers that result in higher breakpoints are considered higher quality, relative to those that result in lower breakpoints. Progressive ratio schedules have been used to identify preferred stimuli for children and adolescents with developmental disabilities (e.g., DeLeon, Frank, Gregory, & Allman, 2009; Francisco, Borrero, & Sy, 2008). Roane, Lerman, and Vorndran (2001) demonstrated a link between breakpoints and reinforcer effectiveness for adolescents with developmental disabilities. In this study, for three participants, the stimuli that produced higher breakpoints were more effective, relative to those that produced lower breakpoints, when used as reinforcers in three different kinds of interventions to reduce destructive behavior.

More recently, researchers have used progressive ratio assessments not to assess reinforcer quality, but to identify individual differences in breakpoints on average across groups. In these cases, a fixed reinforcer is used in the progressive ratio schedule, and the breakpoint is interpreted as a person's tendency to persist in responding as response requirements increase. This is often considered reflective of the trait variable *motivation.* In one such study, Hershenberg and colleagues (2016) found significantly lower breakpoints on average in adults with both unipolar and bipolar depression compared to non-depressed controls. Similarly, in a study of amotivation in adults with schizophrenia, Wolf and colleagues (2014) found significantly lower breakpoints on average for those with schizophrenia compared to controls without any personal or first-degree family experience with psychosis. In both cases, authors hypothesized that an impairment in the functioning of neurological reward systems was the underlying cause of the observed behavioral differences. Chelonis, Gravelin, and Paule (2011) examined breakpoints under a progressive ratio schedule for 847 children from 4- to 14-years-old. They found a systematic increase in breakpoint as a function of age, which they interpreted consistent with a construct of motivation that increases within-individual with development.

If breakpoints do reflect a stable, trait-like tendency to behave, they could be useful to design classroom interventions. For students with low breakpoints, even a low-quality reinforcer that can be delivered following only a few responses might be more effective than a higher-quality reinforcer delivered after many responses. Students with higher breakpoints might respond best to the delivery of a high-quality reinforcer on a lean schedule. Breakpoints might also be useful moderators that predict student response to packaged interventions. Students with high breakpoints might benefit most from an intervention like Check-in Check-out, which involves the delivery of rewards after a daily or weekly goal is met. On the other hand, students with low breakpoints might respond best to a token economy with more frequent reinforcer delivery. Because implementing a rich reinforcement schedule is more effortful than a lean one, school teams could maximize the efficiency of behavioral interventions by only using rich schedules when indicated. Thus, understanding the influence of rate of reinforcement on student behavior might help school teams design interventions to promote appropriate behavior.

However, we do not know the extent to which progressive ratio breakpoints are stable within individual children. That is, we do not know how much estimates of breakpoints could change from test to retest, due to random error or from error related to the assessment day. We also do not know what contributions other kinds of error make to the breakpoint score. It is possible that elements of the assessment procedure (e.g., the type of response required) impact the breakpoint obtained in that assessment. In previous studies using progressive ratio assessments, target responses have included arbitrary responses such as mouse clicks (e.g., Miras et al., 2012), button presses (e.g., Strauss, Whearty, Morra, Sullivan, Ossenfort, &

Frost, 2016), and functional and gross motor responses (e.g., McLeod & Griffiths, 1983; Tiger, Toussaint, & Roath, 2010), but the effects of using one response over another on assessment outcomes are unknown. Before we can evaluate the validity of progressive ratio assessments for predicting or explaining classroom behavior, we must establish breakpoint reliability.

If the contribution of both day-level factors and response-type-level factors to a child's breakpoint can be estimated, researchers can identify assessment protocols that would generate acceptably reliable outcomes (Yoder, Lloyd, & Symons, 2018). For instance, if response-type-level factors are a major source of variance in breakpoint scores, breakpoints determined in assessments with different response types could be averaged to produce a sufficiently stable breakpoint estimate. Likewise, breakpoints determined on different days could be averaged for the same purpose if breakpoints vary substantially day-to-day within individuals. Before the validity of progressive ratio assessments for predicting response to interventions and for individualizing school-based interventions can be assessed, an assessment protocol with adequate reliability is needed.

Delay Discounting Assessments

Delay discounting is the tendency to assign less subjective value to a reward when there is a delay to its receipt (Ainsley, 1974). Though most people discount a delayed reward to some extent, individual differences have been documented in how much value is lost over delays (Odum, 2011b). The degree to which delayed rewards are discounted can be observed in a preference for a smaller reward delivered sooner over a larger reward delivered later. A preference for smaller rewards delivered sooner can be interpreted as evidence of *impulsivity*; it indicates that temporally immediate outcomes have a greater influence on behavior than temporally distant ones (Madden & Johnson, 2010).

As an example, imagine being offered the choice between receiving $100 today or $100 dollars next week. Most people would choose to receive the money sooner, if the amounts were equal. If the choice was between receiving $50 today or $100 dollars next week, responses could differ. People who choose the smaller sooner reward ($50) give an indication that the larger reward ($100) has lost some value due to the week-long wait for its receipt. That is, for these people, $100 a week from now has a subjective value of $50 or less. For those who choose the larger later reward, $100 a week from now retains more than $50 of value despite the delay. The subjective value of a delayed reward can be estimated by determining the point at which an individual changes their preference between the delayed and immediate reward. For example, someone might choose to get $50 immediately over $100 dollars in a week, but would choose to get $100 in a week over $49 dollars immediately. For that person, $100 delayed by a week is worth more than $49, but less than $50, and can be estimated to lie halfway between those amounts at $49.50.

An individual's rate of discounting can be described by estimating the subjective value of a delayed reward at delays of several different lengths. Often, these values are estimated by offering different hypothetical choices between some money now, and a larger amount of money at specific delays (cf. Monetary Choice Questionnaire [MCQ]; Kirby & Marakovic, 1996; Kirby, Petry, & Bickel, 1999). Nonlinear regression can then be used to estimate parameters in a hyperbolic function that best fits those values. A frequently used hyperbolic discounting model is $V = \frac{A}{(1+k \bullet D)}$ (Mazur, 1987) which describes the subjective value (V) of a reinforcer with a current value (A) at some delay (D) in terms of the rate of discounting (*k)*. Here, where D is zero, $V = \frac{A}{1}$ and A retains its full value. As D increases, V approaches zero. The estimated parameter *k* is an index of discounting where greater values reflect greater discounting, as V diminishes faster at any delay (Yoon & Higgins, 2008).

In adults, discounting rates have been predictive of addictive behaviors, including drug or alcohol use, gambling, and other risky behaviors (Kollins, 2003; Reynolds, 2006; Yi, Mitchell, & Bickel, 2010). In adolescents, discounting rates have predicted response to drug treatment programs (Stanger et al., 2012), end-of-year grades (Duckworth & Seligman, 2005), and the development of a smoking habit (Audrain-McGovern et al., 2009). These long-term predictions suggest a stable, trait-like functioning of discounting rate. Some evidence suggests that discounting could be trait-like in younger children as well. Scholars have found that discounting rate predicts response to obesity treatment in young children (Best et al., 2012), and is different, on average, in children with a family history of substance abuse compared to those without such a family history (Dougherty et al., 2014). Further, individual differences in discounting rate might explain a child's tendency to engage productively in the classroom. Reed and Martens (2011) demonstrated correspondence between discounting rate and on-task responding when rewards for being on-task were delayed or immediate. That is, in their sample of 6^{th} grade students, delayed rewards were less effective for those with higher rates of discounting, relative to those with lower rates of discounting.

If discounting rate is reflective of a stable tendency to behave in children, it could function as another way for school staff to select or individualize behavior interventions for students. For some students with high discounting rates, it might be important to prioritize the immediate delivery of reinforcement as part of a behavior plan. For other students, reinforcers might retain value over long delays, which could help school staff when prioritizing limited resources. As suggested by Reed and Martens (2011), discounting might be a moderator that predicts response to certain classroom interventions. Finally, discounting might function in children (as it does in adults) as a predictor of future problematic behavior, and allow for the allocation of early intervention services (Staubitz, Lloyd, & Reed, 2018).

Delay Discounting in Children

Despite this promise, little research to date has examined delay discounting in young children (see Staubitz et al. [2018] for a review). One reason for the relative scarcity of research on delay discounting in children could be that adaptations to the protocols used with adults are needed (Staubitz et al., 2018). Children might have a more difficult time than adults accurately responding to hypothetical questions about their preferences. For adults, discounting of real and hypothetical rewards have a strong correlation (e.g., Johnson & Bickel, 2002; Madden, Begotka, Raiff, & Kastern, 2003), but few studies have evaluated this correspondence for children. In a recent study, Miller (2019) observed similar rates of discounting in 7- to 10-year-old children when rewards were real and hypothetical. However, the sample (n = 9) was too small to establish statistical equivalence between the two kinds of assessments. Examining the effects of stimulant medication on rates of delay discounting in 9- to 12-year-old children with ADHD, Shiels and colleagues (2009) found that methylphenidate reduced discounting, relative to placebo, only when rewards were delivered but not when they were hypothetical. This outcome suggests that reward status (i.e., real vs. hypothetical) has an effect on the measured discounting rate under some conditions. More research is needed on the extent to which discounting rates are correlated across real and hypothetical rewards for young children.

Similarly, for young children who have less experience with money than adults, standard protocols with hypothetical questions about money like the MCQ might not produce accurate results. Other commodities might be needed to assess discounting in children. In adults, the type of commodity used in the assessment has been shown to affect estimates of discounting—with money discounted less than other commodities (e.g., Baker, Johnson, & Bickel, 2003; Koffarnus & Bickel, 2014; Weatherly, Terrell, & Derenne, 2010). However, no studies have evaluated the effects of commodity on the estimates of discounting for children, and few studies have used commodities other than money (Staubitz et al., 2018). Finally, young children could have a difficult time responding to the number of questions required to accurately define their discounting curves. Most discounting assessments are based on a large number of choices, and can be tedious, especially for those with limited attention spans (Koffarnus & Bickel, 2014). Efficient methods for determining discounting rates might be needed for young children in schools.

5-Trial Adjusting Delay Task

The 5-trial adjusting delay task (ADT-5; Koffarnus & Bickel, 2014) allows discounting rate k to be estimated from only five questions. The ADT-5 is possible because k can be calculated mathematically if the length of the delay at which the reinforcer loses half of its value is known. Yoon and Higgins (2008) provide the following proof. Consider the hyperbolic discounting model:

$$V = \frac{A}{(1 + k \bullet D)}$$

At some delay the value of the reinforcer will be reduced by half ($V = \frac{A}{2}$). Yoon and Higgins (2008) refer to this delay as ED_{50}, after a similar term used in pharmacology. Substituting these terms into the hyperbolic discounting model produces the following formula:

$$\frac{A}{2} = \frac{A}{(1 + k \bullet ED_{50})}$$

Cross multiplying and cancelling out A values results in:

$$1 + k \bullet ED_{50} = 2$$

Finally, subtracting one from each side and dividing by *k* produces:

$$ED_{50} = \frac{1}{k}$$

The ADT-5 is designed to determine ED_{50} by asking for a preference between a smaller reward available immediately and twice that reward at some delay, over five trials. For each successive trial, delays are adjusted, to be either shorter (if the immediate reward was chosen) or longer (if the delayed reward was chosen). Koffarnus and Bickle (2014) presented delays from 1 h to 25 years, evenly spaced along a logarithmic scale in their initial evaluation of the ADT-5. In that study, the authors validated the ADT-5 by demonstrating that it identified discounting rates that were correlated with *k* values estimated from the MCQ, and that it was sensitive to manipulations that have known effects on discounting rates (i.e., manipulations of reward status [real vs. hypothetical], amount, and question phrasing) for undergraduate participants.

Miller (2019) developed an adapted ADT-5 to be suited for use with young children. His primary adaptation was using delays arranged on a logarithmic scale from 2 s to 2 h 5 min. These shorter delays might be more comprehensible for young children than delays measured in years. Further, if non-hypothetical rewards are used, the assessment can still be completed in a short time. Miller designed this adaptation so that an entire assessment could be completed in a single 6-h day in a partial hospitalization program. The efficiency of this assessment has promise for use in schools. However, Miller's smaller range of delays might have introduced ceiling effects into this assessment—several children obtained the maximum possible ED_{50} value in his study. It is unknown if these ceiling effects would replicate in a larger sample or with different commodities. Additionally, the study included only nine participants, which was not enough to statistically evaluate questions of stability within-individual and across reward status. Further study of the adapted ADT-5 in a larger sample of young children is needed.

Rationale

In summary, both progressive ratio and delay discounting assessments could have utility to inform behavioral intervention for young children in schools. Progressive ratio assessments have promise for identifying the effects of response effort and reinforcement rate on the behavior of children. Understanding these effects could help school staff choose interventions that are likely to be effective for individual children. Likewise, delay discounting assessments might help school staff individualize treatments and predict responses to interventions. Further, delay discounting assessments might be useful for predicting a child's risk of developing impulsive behavior problems in the future.

However, before validating these assessments for making determinations about treatments, we must evaluate the reliability of assessment results, and identify the non-trait factors that affect scores. Specifically, the within-person stability of breakpoints derived from progressive ratio assessments is unknown and merits investigation. Additionally, an estimation of the contribution of day-related and task-related variance to the observed breakpoints is needed to identify an assessment protocol that produces adequately reliable breakpoints in young children.

For delay discounting assessments, variations of procedures that are adapted to young children require validation. One question is whether such adapted assessments are sensitive to individual differences, or if that sensitivity is limited by ceiling effects. That is, if discounting is assessed, will different rates be identified across different individuals, arranged in some distribution? Only if discounting rates show meaningful variance across individuals can they be useful assessments. Another question relates to the effects that reward status (real vs. hypothetical) and commodity (monetary vs. non-monetary) have on obtained discounting rates. Understanding these effects can inform development of an efficient assessment protocol that produces reliable and sensitive estimates of discounting rate.

If these assessments meet a threshold of stability and sensitivity, additional exploratory questions about the scores they produce are of interest. For example, are either progressive ratio breakpoints or rates of delay discounting correlated with teacher-reported classroom problem behavior? To what extent are breakpoints and discounting rates measuring similar constructs? Investigating these exploratory questions can inform future studies examining the construct validity of progressive ratio and delay discounting assessments.

Purpose

The purpose of this study was to pilot two assessments with young children: a progressive ratio assessment evaluated across two tasks and Miller's adapted ADT-5 delay discounting assessment using varied reward and commodity types. I recruited a sample of young children with and without a history of classroom problem behavior to evaluate the relationship between outcomes of either assessment and histories of problem behavior. This study addressed two primary and three exploratory research questions.

Primary Research Questions

Research Question (RQ) 1: In a sample of elementary-age children with and without problem behavior, what proportion of the variance in progressive ratio scores is due to true score variance, and how much is due to session or task factors? How many sessions and tasks would need to be averaged to yield a progressive ratio score that produces an adequately dependable ranking of children (i.e., a g coefficient greater than .70)?

RQ 2: In a sample of elementary-age children with and without problem behavior, what is the range and standard deviation of k as determined by the adapted ADT-5 using real, non-monetary rewards? What is the range and standard deviation of k as determined by the adapted ADT-5 using hypothetical monetary and non-monetary rewards? Are k values positively correlated in individuals across adapted ADT-5 assessment types?

Exploratory Research Questions

RQ 3: Do children with problem behavior reach a lower breakpoint in a progressive ratio assessment than children without problem behavior? Does this difference depend on the task used in the assessment?

RQ 4: Do children with problem behavior have higher rates of delay discounting than children without problem behavior? Does this difference depend on the commodity used in the assessment?

RQ 5: Is there a non-zero correlation between progressive ratio breakpoints and delay discounting rates?

CHAPTER 2

METHOD

Inclusion and Exclusion Criteria

I recruited two groups of students: students with problem behavior, and students without problem behavior. Students were recruited to the problem behavior group if (a) they were in 1st to 3rd grade, (b) they attended their neighborhood school and were not receiving special education services, (c) their teacher reported frequent problem behaviors in the classroom, and (d) they had mastered single-digit basic math facts. Students were recruited to the non-problem behavior group if (a) they were in the same classroom as a student in the problem behavior group, (b) they were not receiving special education services, (c) they were matched on age, socio-economic status (SES; estimated by eligibility for free or reduced lunch), gender, and teacher-reported math skill with a student in the problem behavior group, and (d) their teacher reported an absence of problem behavior in the classroom.

Students were excluded from any group in this study if they did not speak English at school. They were also excluded from the delay discounting portion of the study if they were unable to answer questions comparing lengths of delays at pretest, or if they indicated a consistent preference for waiting when immediate or delayed quantities were equal. Students were excluded from the progressive ratio blocks assessment if they lacked the fine motor skills needed to put small cubes through a square hole, and were excluded from the progressive ratio math assessment if they did or could not write correct responses to math problems (see pre-test procedures below). Sources of information are listed by inclusion and exclusion criteria for each group in Table 1.

Table 1

Inclusion and Exclusion Criteria and Information Source by Group

Group	Inclusion/Exclusion Criteria	Information Source
Problem Behavior Group	1st to 3rd grade	File review
	No SPED services	File review
	Problem behavior	Teacher report
	Mastered basic math facts	Teacher report; pre-test
	Understand delays	Pre-test
	Can write responses, insert blocks	Pre-test
Non-Problem Behavior Group	Matched age, gender, SES	File review
	Matched math skill	Teacher report
	No SPED services	File review
	No problem behavior	Teacher report
	Mastered basic math facts	Teacher report; pre-test
	Understand delays	Pre-test
	Can write responses, insert blocks	Pre-test

Note. SPED = Special Education; SES = Socio-economic status.

Recruitment

Classroom teachers who were participating in another research project nominated students for participation in this study as part of the problem behavior group. Teachers were asked to identify students who displayed mild-to-moderate disruptive behavior, such as talking to peers during academic activities, out of seat behavior, or calling out in class. Parents of these children were asked to give informed consent for study participation. Once consent was obtained, the classroom teacher indicated whether the student had mastered basic math facts and was likely to meet other inclusion criteria. For each student with problem behavior, the teacher was asked to identify up to four matched peers from the same classroom, and to send home consent forms to their parents. If parents gave their consent and these matched students met inclusion criteria, at least one of them was included in the non-problem behavior group, matched with their classmate with problem behavior.

Participants

A total of 35 students participated in this study. Of these, 32 repeated the progressive ratio assessments three times, two completed them twice and one completed one set of assessments. Two students met inclusion criteria for the arbitrary progressive ratio task, but not the academic one. These students completed the arbitrary task only. A subset of 24 students met inclusion criteria for the delay discounting measure, and all 24 completed all three delay discounting assessments. Participant characteristics are listed by group in Table 2. Students were between 6 years and 5 months and 9 years and 8 months old (*M* = 7 years and 7 months). A substantial portion of the sample (*n* = 15 [43%]) was eligible for free or reduced lunch, and a majority (*n* = 25 [71%]) were boys. Relative to the group of students with teacher-reported problem behavior, the comparison group was significantly younger, $t(33) = -2.231$, $p < .05$, and contained fewer students eligible for free or reduced lunch, $X^2(1, N = 35) = 4.375$, $p < .05$. The subset of participants who completed the delay discounting assessments was similar in age and demographics to the whole sample.

Table 2

Characteristics of Included Participants by Subgroup and Assessment Type

Group	*n*	Mean Age (SD)	FRL Eligible (%)	Boys (%)	White (%)	Black (%)	Latinx (%)	Asian (%)	2 or More Races (%)
Problem Behavior	21	7 y 10 m (10 m)	12 (57%)	14 (67%)	10 (48%)	10 (48%)	--	--	1 (5%)
Non-Problem Behavior	14	7 y 2 m (10 m)	3 (21%)	11 (79%)	8 (57%)	3 (21%)	2 (14%)	1 (7%)	--
Total	35	7 y 7 m (11 m)	15 (43%)	25 (71%)	18 (51%)	13 (37%)	2 (6%)	1 (3%)	1 (3%)
Group	DD *n*								
Problem Behavior	15	8 y 1 m (11 m)	9 (80%)	10 (67%)	6 (40%)	8 (53%)	--	--	1 (7%)
Non-Problem Behavior	9	7 y 3 m (13 m)	3 (33%)	7 (78%)	6 (67%)	2 (22%)	1 (11%)	--	0
Total	24	7 y 9 m (13 m)	12 (50%)	17 (71%)	12 (50%)	10 (42%)	1 (4%)	--	1 (4%)

Note. SD = Standard deviation; FRL = Free or Reduced Lunch; DD = Delay Discounting; y = years; m = months.

Setting

Participants attended one of four public elementary schools located in a large city in the southern United States. All schools had received school-wide training from the local school district in Positive Behavior Interventions and Supports (PBIS) and/or Social and Emotional Learning (SEL) programs, and all had implemented universal (Tier 1) supports for student behavior. School demographics are shown in Table 3. Assessments were conducted one-on-one at tables in quiet hallways or in school libraries. Sessions were scheduled according to teacher preference and student availability. Repeated sessions were conducted at the same time of day and in the same setting as the first session for each student, with four exceptions. For two students, session times were changed from the morning to the afternoon following the first session, and for two students, session locations were changed from a table in the hallway to a table in the library following the first session.

Table 3

Selected Demographics of School Settings

Demographics	School 1	School 2	School 3	School 4
Title I status	Yes	Yes	Yes	No
Race/Ethnicity				
% White	44.8	11.6	25.3	47.4
% Black	40.8	15.3	61.4	42.9
% Hispanic	5.5	66.2	10.2	5.3
% Asian	8.6	6.7	2.4	4.0
% Native/PI	0.0	0.0	0.0	0.0

Note. PI = Pacific Islander.

Materials

For the progressive ratio arbitrary task, materials included 200 2-cm multi-colored plastic cubes, a 19.5 x 33 cm clear plastic shoebox with a 2.1 x 2.1 cm hole cut into the lid, a mat with a grid of 2 x 2 cm squares printed on it, a 32.5 x 37.5 cm collapsible fabric bin to hold the cubes, and an electronic device with a camera (e.g., iPad or cell phone). For the progressive ratio academic task, materials included up to 35 single digit addition worksheets with 30 vertically aligned problems per page. These worksheets were randomly generated from www.mathfactcafe.com/worksheet/buildit according to the following specifications: 6 rows of 5 problems printed in large type with addends from 0 to 10 and correct sums from 0 to 10 (see Appendix A for an example). Both tasks required a variety of small edibles (e.g., M&Ms, Goldfish crackers, fruit snacks), a stop watch (or an electronic device with such a function), a paper and pencil data collection sheet, and clipboard. For all delay discounting tasks, materials included a paper and pencil data collection sheet (see Appendix B) and script (see Appendix C). The real delay discounting task also required a prize box filled with a variety of small school-appropriate tangible rewards (e.g., pen toppers, figurines, stickers, small balls, slime). In a subset of sessions, we used a video camera on a tripod to record sessions.

Assessment Schedule

Each student first completed a pre-test to ensure that they met inclusion criteria. For those that met inclusion criteria, I scheduled up to three assessment sessions. Assessment sessions began immediately following pre-test for some students (n = 21), and on a subsequent day for others (n = 14). Regular assessment sessions were scheduled at least 6 and not more than 21 days apart (*M* = 7.9). At each of the next three sessions, students completed two progressive ratio assessments, one with an arbitrary task and one with an academic task, in a randomly determined order. If the student met inclusion criteria for the delay discounting assessments, they also responded to discounting trials during these sessions. A randomly selected half of these students (n = 12) began with the real, non-monetary delay discounting assessment, and ended with the two hypothetical delay discounting assessments. The other half began with the two hypothetical delay discounting assessments and ended with the real delay discounting assessment. Depending on the delays chosen and the times at which sessions are scheduled, the real delay discounting assessment was not always completed in one session. In these cases, the assessment was continued at the next scheduled session. Both hypothetical assessments were conducted in the same session. When both progressive ratio and delay discounting assessments were scheduled in the same session, we started delay discounting assessments first.

Procedures

File Review and Pre-Test

Following nomination and consent, a research assistant reviewed the student's file and/or interviewed the teacher to confirm that students met inclusion criteria and collected demographic. I also asked teachers to complete the Social, Academic, and Emotional Behavioral Risk Screener (SAEBRS; Kilgus, Chafouleas, Riley-Tillman, & von der Embse, 2014) at this time. Research assistants then met one-on-one with the student to complete pre-test activities to confirm final inclusion criteria. At pre-test, students completed a sample worksheet of 10 single-digit addition problems (i.e., one third of the problems shown in Appendix A), and were asked to put small cubes through a square hole. Students who completed the math problems with 90% accuracy and fit a cube through the hole within 5 s of a research assistant's prompt to do so met inclusion criteria for the academic and arbitrary progressive ratio assessments, respectively.

Research assistants also probed students' receptive understanding of delays and time units. First, they presented the time units used to describe delays—seconds, minutes, and hours—with the following script:

> A second is a short amount of time. It takes one second to say your name. It takes about 15 seconds to say the Pledge of Allegiance. A minute is a little longer. It takes about one minute to wash your hands. It takes 30 minutes to watch a TV show. One hour is even longer. It takes one hour to watch two TV shows.

Research assistants then asked students the following questions:

1. Which is longer, one second or one minute?
2. Which is longer, one minute or one hour?
3. Which is longer, one hour or one second?
4. Would you rather have a piece of candy now, or wait one minute to have a piece of candy?
5. Would you rather play a game now, or wait one hour to play a game?

6. Would you rather get a prize now, or wait five seconds to get a prize?

Students who answered Questions 1–3 correctly and indicated a preference for the immediate reward in at least two of Questions 4–6 met inclusion criteria for the delay discounting assessments. Students who answered any of the first three questions incorrectly (n = 4) were excluded from the delay discounting assessments. If a student indicated a preference for a delayed reward in more than one of the hypothetical questions where quantities were equal, I probed their preferences with two similar questions about real rewards. A research assistant asked:

1. I have a [edible treat] here. You can have one. Do you want it right now, or do you want to wait 30 seconds before you have it?
2. I have a prize box here. You can choose a prize from it. Do you want to get a prize right now, or do you want to wait 1 minute before you get it?

If the student requested the item right away, the research assistant granted immediate access and asked the next question. If the student requested the item after a delay, the research assistant set a timer, removed the timer from view, delivered the item at the end of the interval, and then asked the next question.

Students who expressed a preference for the delayed reward in both of these questions (n = 7) were excluded from the delay discounting assessment. I interpreted this pattern of responding to indicate either that students did not understand the question or that some factor other than the quantity of the reward (likely the social desirability of the response) was the primary influence on their choice. I included students who chose the immediate reward for at least one question at this stage.

Progressive Ratio

In the progressive ratio assessment, students earned small edible rewards for completing increasing schedule requirements of a brief task response. Schedules increased within sessions in a rapid additive series (Reed et al., 2009). Schedules increased according to the following pattern: fixed ratio 1 (FR-1), FR-2, FR-5, FR-10, FR-20, FR-30, FR-40. Further schedules increased by intervals of 15 for three schedules (FR-55, FR-

70, FR-85). The assessment ended when the student expressed a desire to stop or if 1 min elapsed without a response. If neither of these requirements was met after 10 min, the assessment was ended after the student received the next reinforcer.

The rapid additive series of increasing schedules was chosen after considering the balance between sensitivity, efficiency, and the risk of participant satiation on the reinforcer. The sensitivity of the progressive ratio assessment is determined by the number of schedules presented to the participant and the difference in number of responses between adjacent schedules. All else equal, adding schedules and decreasing the steps between them would produce a more sensitive measure of the effect of schedule of reinforcement on participants' behavior. However, it would also increase assessment time, and add to the number of reinforcers earned, which could increase the risk that satiation on the reward may affect the results. The rapid additive approach used by Reed and colleagues (2009) offers a shorter assessment with less risk of satiation than a simple arithmetic approach (e.g., where the response requirement is increased by one in each successive schedule; cf. Glover et al., 2008), but a more sensitive measure than geometric approaches where one schedule is multiplied by a constant to determine the next schedule (e.g., when schedule requirement doubles in each successive schedule; cf. Tiger et al., 2010).

We conducted two progressive ratio assessments, one with an arbitrary task, and one with an academic task. For the academic task, students provided written responses to single-digit addition problems printed on worksheets. For the arbitrary task, students placed a plastic cube through a small hole into a plastic box, one at a time. Each math problem solved and each cube inserted counted as one response. At the beginning of each assessment, research assistants offered a choice of edible rewards to be earned in that session, and gave directions for the task by reading the task script (see Appendix D). The scripted directions included a model of the task and single example exposure to the contingency, as well as a reminder to the student that he or she can stop whenever they like by saying "stop" or touching a laminated card depicting a stop sign. During the assessment, research assistants collected data on the number of responses, provided rewards according to the programmed schedule, and otherwise minimized interaction with the student. If participants attempted to put more than one cube in the hole at a time, or hold more than one cube in their hands, research assistants blocked access to the box and reminded students to put the cubes in one at a time. If participants protested during the session, inquired about the schedule, or engaged in problem behavior, research assistants reminded them “You can keep earning snacks, or you can say *stop* if you want to stop. It’s your choice”. I programed responses for research assistants when students refused to engage with the task or asked to terminate sessions, but this never happened with any student.

Delay Discounting

Participants who met delay discounting inclusion criteria completed three delay discounting assessments using the ADT-5 format (Koffarnus & Bickel, 2014) with delays adapted for young children (Miller, 2019). In all delay discounting assessments, students were offered the choice between a larger reward at some delay and a smaller reward immediately over several trials. The reward amounts were held constant and the larger reward was always twice the value of the smaller one. The first trial always began with a delay of 1 m 45 s. At each trial afterwards, the delay changed, with the direction of change determined by the previous choice made. If the immediate reward was selected, the next trial had a shorter delay. If the delayed reward was selected, the next trial had a longer delay. A flowchart representing the response guided changes to delays is shown in Figure 1.

The possible delays were adapted from delays used by Miller (2019), with the following minor adjustments for ease of comprehension: delays longer than 50 s were rounded to the nearest 15-s increment (e.g., 56 s is rounded to 60 s, 100 s is rounded to 105 s) and delays longer than 1 h were rounded to the nearest whole min (e.g., 1 h 33 min 45 s [5625 s] was rounded to 1 h 34 min [5640 s]). Delays were arranged to be evenly spaced on a logarithmic scale from 2 s to 2 h 5 min. Changes in delay were programmed so that the first adjustment is up or down approximately one logarithmic unit (i.e., from 105 s to 1005 s or from 105 s to 10 s), with each successive adjustment decreasing by half the previous step in log units (Miller, 2019, p. 50). As described by Miller (2019), if the immediate choice was made on every trial, only four trials were

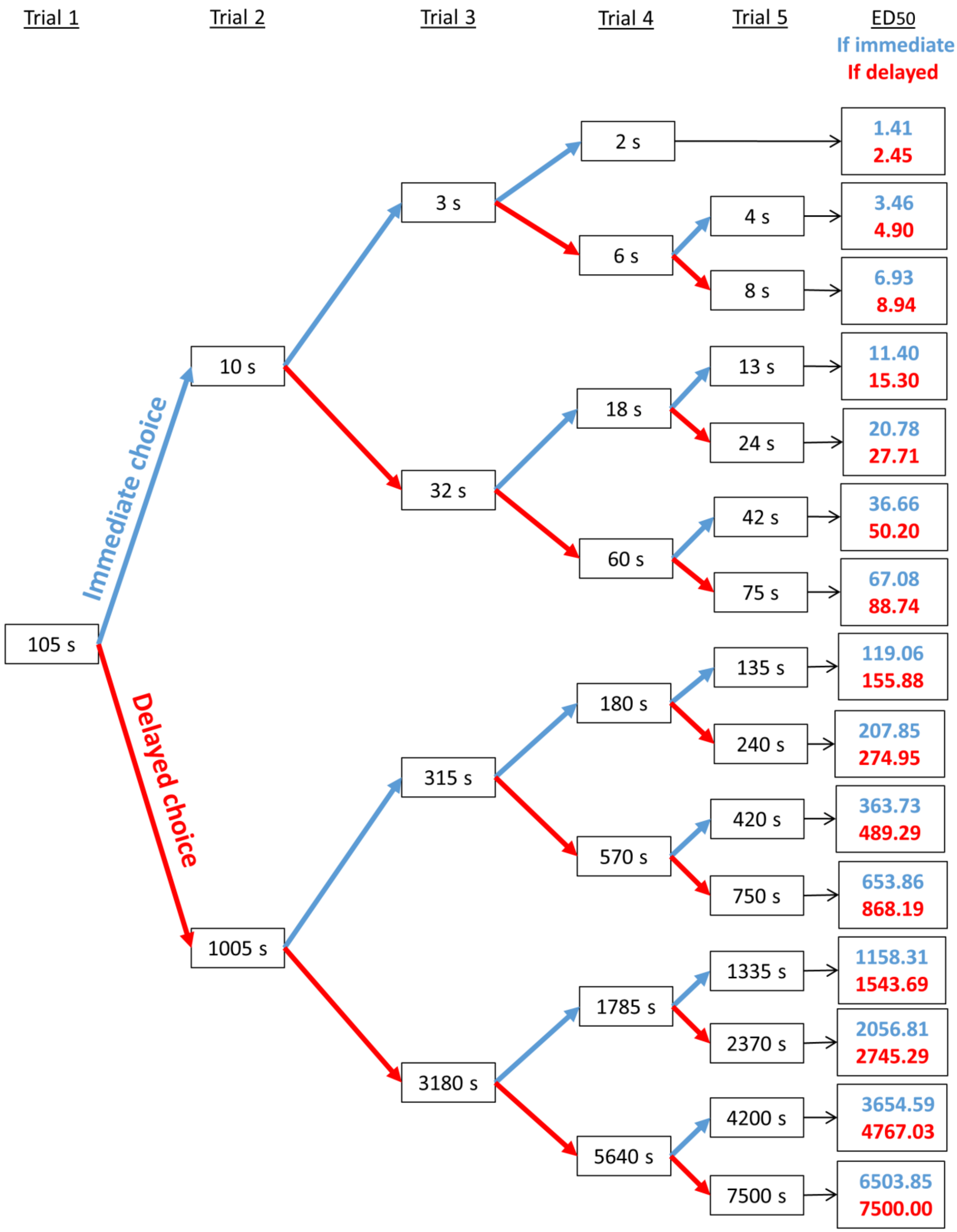


Figure 1. Response-guided changes to delays and outcomes for the ADT-5 delay discounting measure.

completed, because adjustments from a 2 s delay would have required precision less than 1 s. In all other cases, students responded to five trials.

The ADT-5 assessment with real rewards was conducted according to the following procedures. First, the research assistant showed the reward box to the participant and told them they could choose some items from it after they answer some questions. The research assistant provided time anchors by reminding the participant “It takes 1 second to say your name, 15 seconds to say the Pledge, 1 minute to wash your hands, and 30 minutes to watch a TV show.” and then solicited and answered any questions. Then, the research assistant presented the first trial by saying “Do you want to pick one treat from the treasure box right now OR do you want to pick two treats from the treasure box in 1 minute and 45 seconds?” If the student chose the immediate reward, the research assistant allowed the participant to choose one item from the treasure box, and moved on to the next trial. If the student chose the delayed reward, the research assistant set a timer for the delay interval and told the participant “This timer will remind me when to get the treasure box for you so you can pick your treats”. The timer was not displayed throughout the interval. When the delay elapsed, the participant was allowed to choose two items from the treasure box. After the two items were chosen, the next trial was offered. All delays were described in conventional seconds, minutes, and/or hour units; that is, 60 s was described as “1 minute” and 750 s as “12 minutes and 30 seconds”. Conversions between total seconds and conventional time units are shown in Table 4.

Time in the delay interval was unprogrammed and activities within the interval varied across participants. During the interval, the research assistant did not provide high quality attention or access to usually-restricted activities like access to the playground, gym, or computers, but answered questions and permitted access to quiet independent activities (e.g., books, quiet drawing). If delays were long enough to accommodate the completion of other scheduled tasks, research assistants presented other scheduled assessments during the delays. For delays over 20 min, research assistants escorted participants back to their classrooms, and returned to deliver the reward when the delay elapsed. Before presenting a trial, the research

Table 4

Trials, Delays, and Outcomes for Delay Discounting Assessments

		Delay	Delay	ED50 (s) if last choice is		*k* if last choice is:		Total prizes if last choice is:		Total delay if last choice is:	
Index	Trial	(s)	(s, min, h)	Immediate	Delayed	Immediate	Delayed	Immediate	Delayed	Immediate	Delayed
1	4	2	2 s	1.41	2.45	0.707107	0.408248	4	5	0	2
2	3	3	3 s								
3	5	4	4 s	3.46	4.90	0.288675	0.204124	6	7	3	7
4	4	6	6 s								
5	5	8	8 s	6.93	8.94	0.144338	0.111803	7	8	9	17
6	2	10	10 s								
7	5	13	13 s	11.40	15.30	0.087706	0.065372	6	7	10	23
8	4	18	18 s								
9	5	24	24 s	20.78	27.71	0.048113	0.036084	7	8	28	52
10	3	30	30 s								
11	5	42	42 s	36.66	50.20	0.027277	0.01992	7	8	42	84
12	4	60	1 min								
13	5	75	1 min 15 s	67.08	88.74	0.014907	0.011269	8	9	102	177
14	1	105	1 min 45 s								
15	5	135	2 min 15 s	119.06	155.88	0.008399	0.006415	6	7	105	240
16	4	180	3 min								
17	5	240	4 min	207.85	274.95	0.004811	0.003637	7	8	285	525
18	3	315	5 min 15 s								
19	5	420	7 min	363.73	489.29	0.002749	0.002044	7	8	420	840
20	4	570	9 min 30 s								
21	5	750	12 min 30 s	653.83	868.19	0.001529	0.001152	8	9	990	1740
22	2	1005	16 min 45 s								
23	5	1335	22 min 15 s	1158.31	1543.69	0.000863	0.000648	7	8	1110	2445
24	4	1785	29 min 45 s								
25	5	2370	39 min 30 s	2056.81	2745.29	0.000486	0.000364	8	9	2895	5265
26	3	3180	53 min								
27	5	4200	1 h 10 min	3654.59	4867.03	0.000274	0.000205	8	9	4290	8490
28	4	5640	1 h 34 min								
29	5	7500	2 h 5 min	6503.85	7500.00	0.000154	0.000133	9	10	9930	17430

assistant ensured the remaining time in the school day and her schedule permitted the delivery of the reward at the designated delay. If schedules could not accommodate the delay, the trial was not presented, and the delay discounting assessment continued on another assessment day. Thus, participants who chose the longest delays had their assessments split over several sessions. The total number of treats each participant could receive, and the total delays they could experience, are shown in Table 4.

We also conducted two hypothetical delay discounting assessments, one where the choice was between one quarter (25 cents) immediately and two quarters (50 cents) after the delay, and the other where the choice was between one item from the prize box immediately and two items from the prize box after the delay. Procedures were identical to those described above for presenting trials with two exceptions. First, in the directions, research assistants clarified that all prizes and money were hypothetical in this assessment (e.g., "Today, I'm going to ask you to imagine getting a prize from my treasure box. I'm going to ask you questions about when you'd like to get the prizes, and how many you'd like to get. Today, all the prizes will be imaginary."). Second, all trials were presented one after the other with no delays between. Both hypothetical assessments were completed on the same day in a randomly determined order.

If a student asked a question about the length of the delay after a trial was presented, the research assistant repeated the time anchors (i.e., "it takes 1 second to say your name, 15 seconds to say the Pledge, 1 minute to wash your hands, and 30 minutes to watch a TV show"). If a student did not make a choice within 5 s of the presentation of the trial, the research assistant prompted the choice again. I programed research assistant responses if a student refused the task or engaged in problem behavior, but this never happened.

Research Assistant Training

Research assistants were trained according to the following schedule. First, they attended a 1-h meeting, where I presented an overview of these procedures and provided session scripts. Next, research assistants role-played assessment procedures with corrective feedback until they completed each assessment with 90% accuracy, as measured by live procedural fidelity data collection. Then, research assistants practiced using the primary and procedural fidelity data collection forms for the delay discounting assessment during role-play. They repeated sessions until implementing research assistants agreed with an observing coder at 90% over two consecutive role-plays. Finally, research assistants practiced collecting primary and procedural fidelity data from a set of videos showing similar progressive ratio procedures until they agreed with a master code at 90% for two consecutive videos.

Though I planned corrective trainings if procedural fidelity or inter-observer agreement fell below certain benchmarks, agreement and fidelity remained high and consistent across assessments so these corrective trainings were not implemented.

Measures and Outcomes

Classroom Problem Behavior

The SAEBRS is a behavioral risk screener with three subscales for social, academic, and emotional behavior. Scores in these subscales are summed to determine a total behavior score. For elementary aged students, internal consistency reliability exceeds .80 for all subscales (Cronbach's alpha range, .83-.93) with moderate-to-high diagnostic accuracy for at-risk status and high and significant concurrent validity with the Behavioral and Emotional Screening System (Kilgus et al., 2016). The SAEBRS is scored such that low numbers indicate problems and high numbers indicate competencies. A total behavior score of 36 or below indicates at risk status.

Progressive Ratio

Performance on the progressive ratio assessment was measured by tallying the number of tasks completed in each session. Completion of the arbitrary task was defined as the cube completely crossing the plane of the lid through the square hole. Completion of the academic task was defined as the legible printing of a digit (whether correct or incorrect) underneath the math problem on the provided worksheet. Incorrect answers were accepted to reduce the influence of the tendency to make errors when completing mastered tasks on the outcomes. Though inclusion criteria should have ensured that participants had mastered basic math facts, students still may make errors due to inattention or working memory deficits, and this tendency may not be equally distributed across groups. For both progressive ratio tasks, primary data were collected live via event recording (i.e., tallies on the data collection sheets). Assessment results were summarized as (a) the total number of responses completed in each assessment, and (b) the breakpoint, which is the value of the largest reward schedule completed in each assessment.

Delay Discounting

Research assistants used live recording to collect trial-based data on student selection of the larger, later reward or the smaller, sooner reward (see Appendix B). The choice of immediate or delayed reward on the final trial determined the ED_{50}, which is the delay at which the present value of the reward is equal to half the value of the delayed reward. The ED_{50} is calculated by finding the geometric mean of the final delay and the last delay at which the opposite choice was made. That is, if the participant chose the immediate reward at 4 s in Trial 5, the ED_{50} was the geometric mean of 4 s and 3 s (i.e., the delay at which the participant last chose the delayed reward) or 3.46. If the participant chose the delayed reward at 8 s in Trial 5, the ED_{50} was the geometric mean of 8 s and 10 s (i.e., the delay at which the participant last chose the immediate reward), or 8.94. Because $ED_{50} = \frac{1}{k}$ (Yoon & Higgins, 2008), each ED_{50} value was converted into *k* by taking the inverse. All possible delays, ED_{50} values, and *k* values are shown in Table 4.

Inter-Observer Agreement

I evaluated inter-observer agreement (IOA) on 85% of progressive ratio sessions (72% of arbitrary tasks sessions, and 94% of academic task sessions) by permanent product, distributed across participants and research assistants. For the academic task, research assistants collected the completed worksheets, and a second data collector tallied the number of completed problems on the worksheets. For the arbitrary task, after the assessment, the implementing research assistant emptied the box, arranged the cubes on a mat with a grid pattern printed on it, and photographed the cubes. A second data collector tallied the number of cubes that were in the picture. I compared these tallies to the number expected for that session based on the primary data collector's tallies. In both cases, I calculated total agreement as a formative measure by dividing the smaller tally by the larger tally and multiplying by 100. I programmed re-training activities if the average IOA fell below 90%, however, all sessions were above 93%, and these re-training activities were never needed.

As a summative measure of IOA, I calculated intra-class correlations (ICCs) comparing the consistency of the data collected by permanent product and the data collected live. Further, for 13% of progressive ratio sessions that were recorded on video, a second independent data collector tallied responses. I compared these data with the data collected live by implementers by calculating ICCs. Across IOA types, reliability was high. ICCs were greater than .966, and mean total agreement across sessions was above 99% for both assessment types. Reliability indices are listed by data collection method and assessment type in Table 5.

Table 5

Progressive Ratio Assessment Reliability

PR Type	Permanent Product IOA			Video IOA
	Total Mean	Range	ICC	ICC
Academic Task	99.4%	93.4%-100%	.966	.999
Arbitrary Task	99.8%	93.9%-100%	.999	.999

Note. PR = Progressive Ratio; IOA = Inter-observer agreement; ICC = Intra-class Correlation.

I evaluated IOA on 51% of delay discounting sessions by comparing data collected by a second, independent data collector to the implementer's data. I calculated exact agreement on student selection by summing the number of trials with agreement, dividing by the total number of trials, and multiplying by 100. IOA was below 100% in one session, because one coder failed to record a student's response on one trial.

This issue was addressed with retraining, and the discrepancy was resolved. Mean IOA for delay discounting sessions was 99% (range, 80%-100%).

Procedural Fidelity

Fidelity of Assessment Implementation

For progressive ratio assessments, an independent coder assessed procedural fidelity on 50% of sessions by using a checklist. Fidelity coders recorded the presence or absence of four critical scripted pre-assessment behaviors (i.e., offering a choice of snack, modeling the response for the participant, letting the participant practice the response and contact the reinforcer, and describing and presenting the stop card). The absence of any of these behaviors was coded as an error. During the assessment, fidelity coders recorded whether implementers followed the prompting rules, minimized attention, blocked attempts to make multiple simultaneous responses (arbitrary task only), and responded as programmed to problem behavior in each schedule attempted by the participant. If '*no*' was indicated for any item, that was coded as an error. Fidelity coders also indicated whether a reward was delivered at the end of the schedule, which was correct if the participant completed the required number of responses, and an error if the participant completed fewer than the required number of responses. For each session, the number of correct steps was divided by the sum of correct and incorrect steps, and multiplied by 100 to determine the percentage of procedural fidelity for that session. Procedural fidelity exceeded 90% in all sessions. Mean session procedural fidelity was 99% (range, 91%-100%).

As a summative measure of procedural fidelity, I also evaluated the correct implementation of each implementer behavior across sessions by dividing the number of times the behavior was completed correctly by the total number of times the behavior was programmed and multiplying by 100. Across behaviors, correct implementation averaged 99% (range, 87%-100%). The behavior with the lowest fidelity (87%) was blocking an attempt to engage in more than one response at a time during the arbitrary task assessment. One reason for the low percentage was the relative infrequency of attempts to put more than one block into the box at a time. In only 69 of the 379 schedules observed for procedural fidelity did a student attempt to complete more than one response at a time. Sixty of these times, the implementing research assistant responded correctly. The low percentage reflects only nine individual errors. Summative measures of procedural fidelity are listed by method and assessment type in Table 6.

Table 6

Progressive Ratio Procedural Fidelity

PR Type	Across Sessions		Across Behaviors	
	Mean	Range	Mean	Range
Academic Task	99.3%	93.6%-100%	99.6	97.2%-100%
Arbitrary Task	98.6%	90.9%-100%	97.8	86.9%-100%

Note. PR = Progressive Ratio.

For delay discounting assessments, implementing research assistants self-reported procedural fidelity on the delay discounting data collection sheet. For each session, implementing research assistants indicated whether they correctly presented time anchors, solicited and answered questions, and communicated reward type and commodity. For each trial, implementing research assistants recorded whether they offered the correct choice at the correct delay, and whether they responded to in-session behaviors as programmed. A *no* recorded for of any of these behaviors was considered an error. When rewards were non-hypothetical, they also recorded whether they correctly granted access to the treasure box (either immediately or after a delay). The time of access was defined as the time the research assistant granted verbal permission to choose a treat. If the immediate reward was chosen, correct implementation was the research assistant giving verbal permission to choose from the box within 3 s. If the delayed choice was chosen, correct implementation was the research assistant giving verbal permission to choose from the box within 3 s or within 10% of the programmed delay, whichever was greater.

I calculated a summative measure of procedural fidelity by dividing the number of correctly implemented steps in each session by the total number attempted and multiplying by 100. Average procedural fidelity for the delay discounting assessment sessions was 99.6% (range, 92%-100%). I also summarized procedural fidelity by behavior, by dividing the number of times a behavior was implemented correctly by the number of times it was programmed across sessions and multiplying by 100. Average correct implementation of behaviors was 99.6% (range, 97%-100%), including 99% correct reward access in the non-hypothetical assessment.

Inter-Observer Agreement on Fidelity of Assessment Implementation

For progressive ratio assessments, a second coder assessed IOA on procedural fidelity by video on 28% of sessions where procedural fidelity was assessed (14% of total progressive ratio assessments). For delay discounting assessments, a second live coder assessed IOA on self-reported procedural fidelity on 49% of sessions. In both cases, the percentage of agreement was calculated by dividing the number of exact agreements on each implementer behavior by the number of agreements plus disagreements and multiplying by 100. Average IOA on progressive ratio procedural fidelity was 99% (range, 94%-100%). For delay discounting procedural fidelity, IOA averaged 99% (range, 98%-100%).

Data Analysis

I entered all assessment data in electronic spreadsheets, and research assistants checked data entry for accuracy by comparing spreadsheets to paper data collection sheets. If discrepancies were found, research assistants changed the data on the spreadsheet. I checked all discrepancies by comparing the data collection sheets with the changed entry. Unless otherwise noted, statistical calculations were performed in SPSS.

To address RQ 1, I conducted a three-facet generalizability (G) study. First, I used a random effects ANOVA to estimate the proportion of score variance associated with person (σ_p^2), task (σ_t^2), session (σ_s^2), person by task (σ_{pt}^2), person by session (σ_{ps}^2), task by session (σ_{ts}^2) and person by session by task ($\sigma_{pio,e}^2$). I calculated an absolute g coefficient using the following equation (Shavelson & Webb, 2006):

$$g = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_\Delta^2}$$

Where σ_{Δ}^2 is the total error variance, calculated by:

$$\sigma_{\Delta}^2 = \frac{\sigma_t^2}{n_t'} + \frac{\sigma_s^2}{n_s'} + \frac{\sigma_{pt}^2}{n_t'} + \frac{\sigma_{ps}^2}{n_s'} + \frac{\sigma_{ts}^2}{n_t' n_s'} + \frac{\sigma_{pio,e}^2}{n_t' n_s'}$$

Where n_t' is the number of tasks (2) and n_s' is the number of sessions (3). This g coefficient is analogous to a reliability coefficient under classical test theory for decisions made from an individual's absolute score. Then, I conducted a Decision (D) study (Yoder et al., 2018) to evaluate the effects of averaging breakpoints across different numbers of sessions and tasks on overall score dependability (g). I calculated g coefficients with different numbers used for n_t' and n_s' (to represent different hypothetical numbers of sessions and tasks, respectively) in the above equation for total error variance (σ_{Δ}^2). I identified the number of sessions and tasks that would produce g coefficients of .7 or above.

To address RQ2, I constructed figures showing the distributions of *k* values obtained from all three delay discounting assessments (real non-monetary, hypothetical monetary, and hypothetical non-monetary), and visually inspected these figures for evidence of skew that could reflect floor or ceiling effects. Because *k* values were arranged on a logarithmic scale, I log-transformed them and calculated the range, inter-quartile range, and standard deviation of log(*k*)s for each delay discounting assessment type. I also calculated means and 95% confidence intervals for the log(*k*)s from each delay discounting assessment. I calculated Spearman's ρ rank correlations between *k* values obtained from the real and hypothetical non-monetary assessments and between *k* values from the hypothetical monetary and non-monetary assessments, and evaluated their significance with 2-tailed tests.

Then, I examined whether the tests using hypothetical and real prizes, and the tests using hypothetical monetary and non-monetary rewards were equivalent using the two one-sided test (TOST; Lakens, 2017). An inversions of a typical null hypothesis significance test, the TOST is used to evaluate the evidence against the null hypothesis that there **is** a meaningful difference between two means. To use the TOST, researchers first define the smallest difference in means that would be of interest (i.e., the equivalence margin). Then, two *t*-tests are evaluated: one to determine whether there is enough evidence to reject the null hypothesis that there is a positive difference at least as large at the equivalence margin, and another to determine whether there is enough evidence to reject the null hypothesis that there is a negative difference at least as large as the equivalence margin. If both *t* values are significant (i.e., both nulls can be rejected), I would conclude that the two assessments have equivalent results.

Because there is little theory to establish a minimum meaningful difference in *k*, I replicated Miller's (2019) method of setting the equivalence margin. I listed all the possible log(*k*) values that could have been identified by the ADT-5. Then, I determined the difference between each value and the next highest value and calculated the mean of those differences: 0.13. That is, I set the equivalence margin to the average difference detectable by the ADT-5. I used the following equations for TOST of dependent means (Lakens, 2017) to determine whether a difference less than -0.13 (t_L) or greater than 0.13 (t_U) existed between the pairs of means:

$$t_{\mathrm{L}} = \frac{\overline{M_1} - \overline{M_2} - -0.13}{\frac{\sqrt{SD_1^2 + SD_2^2 - 2 * r * SD_1 * SD_2}}{\sqrt{N}}} \quad \text{and} \quad t_{\mathrm{U}} = \frac{\overline{M_1} - \overline{M_2} - 0.13}{\frac{\sqrt{SD_1^2 + SD_2^2 - 2 * r * SD_1 * SD_2}}{\sqrt{N}}}$$

Where $\overline{M_1}$ and $\overline{M_2}$ are the assessments' means, SD_1 and $SD_1$are the assessments' standard deviations, and *r* is the correlation between the two assessments.

To answer RQ3, I tested Pearson's *r* correlations between total SAEBRS scores and the breakpoints obtained from both the academic and arbitrary progressive ratio assessments with two-tailed significance tests against α values of .05 and .01. For students who repeated the assessments more than once, I used the average scores across administrations for these correlations. To determine whether the relationship between problem behavior and breakpoint depended on the task used in the progressive ratio assessment, I compared the SAEBRS and academic breakpoint correlation to the SAEBRS and arbitrary breakpoint correlation. I calculated Steiger's *z* (Steiger, 1980) to test the difference between these two correlations using a web utility (Lee & Preacher, 2013). A significant *z*-score would indicate that different relationships between SAEBRS score and breakpoint were observed when tasks were different.

As another way to examine the relationship between task type, participant characteristics, and responding, I constructed demand curves for each assessment type and participant group. I totaled the number of reinforcers earned at each schedule across participants and sessions, and used the Demand Curve Analyzer (Gilroy, Kaplan, Reed, Koffarnus, & Hantula, 2018) to fit demand curves to these data. The Demand Curve Analyzer models the effects of pricing (schedule of reinforcement) on consumption (reinforcers earned), to identify the point at which an increase in price reduces consumption (P_{MAX}). Larger values for P_{MAX} indicate that responding was not substantially affected by response requirement until higher schedules were reached, while smaller values for P_{MAX} indicate sensitivity to schedule changes at smaller response requirements (Gilroy et al., 2018).

To answer RQ4, I tested Pearson's *r* correlations between total SAEBRS scores and the log(*k)* values from the real prizes, hypothetical prizes and hypothetical money delay discounting assessments with two-tailed significance tests against α values of .05 and .01. To determine whether the relationship between problem behavior and discounting depended on the reward used to assess it, I compared correlations using Steiger's z (Steiger, 1980) as above. A significant *z*-score would indicate that relationships between SAEBRS score and rate of discounting varied depending on reward type.

To answer RQ 5, I calculated a composite variable for progressive ratio scores by computing the mean of all three administrations of both tasks. I calculated a composite of discounting rates by computing the mean of the log(*k*) values from each of the three delay discounting assessments. I tested the Pearson's *r* correlation of these composite variables against an α of .05.

CHAPTER 3

RESULTS

Results of the SAEBRS are shown in Table 7, for the whole sample and for the subgroup included in the delay discounting assessments. Mean scores were lower (indicating higher risk) for the problem behavior group, relative to the non-problem behavior group for total score, all subscales, and the item rating impulsiveness. Means and standard deviations for composite breakpoints from both the arbitrary and academic progressive ratio assessments are shown in Figure 2. On average, higher breakpoints were observed in the arbitrary task (*M* = 42) than the academic task (*M* = 32). In each assessment, mean breakpoint for students in the problem behavior group was slightly higher than for students in the non-problem behavior group, though for all differences, 95% confidence intervals overlapped.

Table 7

Total and Subscale SAEBRS Scores by Participant Groups

			Subscale Scores			
Group	*n*	Total Score (*SD*)	Social (*SD*)	Academic (*SD*)	Emotional (*SD*)	Impulsiveness Item (*SD*)
Problem Behavior	21	27.1 (7.14)	7.2 (3.32)	6.8 (3.32)	12.7 (3.31)	.95 (1.07)
Non-Problem Behavior	14	47.3 (6.46)	15.8 (1.63)	14.0 (3.55)	17.5 (2.85)	2.7 (.47)
Total	35	35.2 (12.11)	10.6 (5.07)	9.7 (4.92)	14.6 (3.94)	1.7 (1.24)
Group	DD *n*					
Problem Behavior	15	25.0 (6.43)	7.2 (3.47)	6.3 (3.56)	11.5 (2.85)	.93 (1.10)
Non-Problem Behavior	9	45.4 (5.70)	15.1 (1.54)	13.2 (3.38)	17.1 (2.67)	2.6 (.53)
Total	24	32.67 (11.78)	10.2 (4.84)	8.9 (4.85)	13.6 (3.88)	1.5 (1.22)

Note. SAEBRS = Social, Academic, and Emotional Behavioral Risk Screener (Kilgus et al., 2014); SD = Standard deviation; DD = Delay Discounting. Lower scores indicate higher risk, A total behavior score of 36 or below indicates “at risk” status.

Sample	Task	Mean	(SD)
Total	Academic	32	(18)
PB	Academic	33	(20)
NPB	Academic	29	(17)
Total	Arbitrary	42	(19)
PB	Arbitrary	43	(17)
NPB	Arbitrary	40	(22)

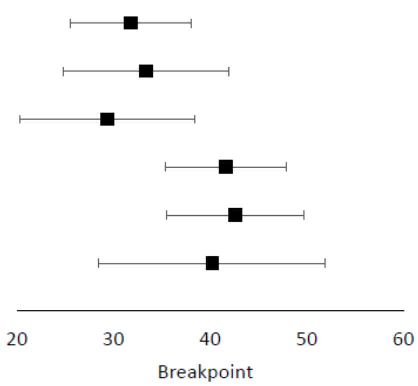


Figure 2. A comparison of means and standard deviations of breakpoints from each type of progressive ratio assessment by participant group. Error bars represent 95% confidence intervals.

Means and standard deviations for the log-transformed *k* values from the ADT-5 with hypothetical non-monetary rewards, hypothetical monetary rewards, and real non-monetary rewards are shown in Figure 3. For all assessments, the problem behavior group had a higher mean discounting rate than the non-problem behavior group, though this difference was smallest in the assessment with real rewards (*M* = -2.6 and - 2.7, respectively). However, 95% confidence intervals overlapped for all assessment and subgroup comparisons.

Sample	Reward	Mean	(SD)
Total	Hy. Prizes	-2.8	(1.0)
PB	Hy. Prizes	-2.6	(1.1)
NPB	Hy. Prizes	-3.1	(0.8)
Total	Hy. Money	-2.5	(1.1)
PB	Hy. Money	-2.4	(1.2)
NPB	Hy. Money	-2.8	(0.4)
Total	Real Prizes	-2.6	(0.8)
PB	Real Prizes	-2.6	(0.7)
NPB	Real Prizes	-2.7	(1.0)

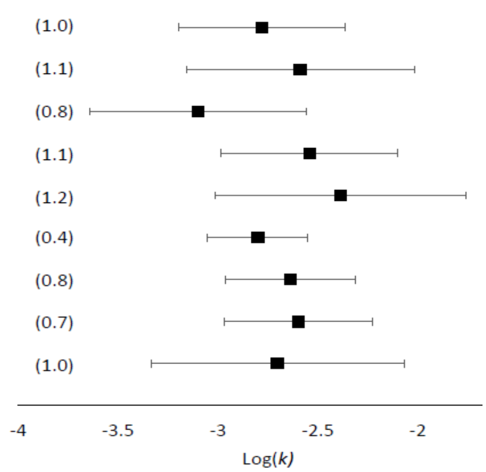


Figure 3. A comparison of means and standard deviations of log(*k*) values from each type of delay discounting assessment by participant group. Error bars represent 95% confidence intervals.

Correlations between measured variables are shown in Table 8. Significant moderate-to-high positive correlations were found between SAEBRS subtests and total scores, and between each subtest ($r(33) = .448$ to .882, $p < .01$). Similarly, significant moderate-to-high positive correlations were found between the two progressive ratio tasks ($r(31) = .542$, $p < .01$) and between pairs of delay discounting assessments ($r(22) =$ .427 to .884, $p < .01$). Visual inspection of scatterplots suggested that relationships between variables were linear and homoscedastic.

Research Question 1: Dependability of Progressive Ratio Scores

The results of the G study are shown in Table 9. The largest proportion of the variance in breakpoint was explained by the object of measurement (person; 35%), though a substantial variance was explained by task (10%) and person x task (23%). This suggests that the task type (i.e., arbitrary vs. academic) had an effect on rankings across children, and that this effect varied by person. Though the session alone accounted for none of the variance in breakpoint, the person x session interaction explained 9% of the variance. Thus, rankings of persons by breakpoint varied more by task than by session. The unexplained variance was non-trivial (21%). The assessment protocol used in this study, where two tasks were repeated in three sessions, was associated with an absolute *g* of .59 , meaning that nearly 60% of the variance in breakpoint scores was accounted for by true score.

Table 9

Percentage of Variance in Progressive Ratio Breakpoints Explained by Each Source

Source	Percentage of variance explained
Person	35%
Session	0%
Task	10%
Person x Session	9%
Person x Task	23%
Session x Task	0%
Person x Session x Task	21%

Table 8

Pearsons' r Correlations Among Study Variables

	Measure	1	2	3	4	5	6	7	8	9	10	11
1.	Age	--										
2.	SAEBRS Total	-.344*	--									
3.	SAEBRS Social	-.361*	.878**	--								
4.	SAEBRS Academic	-.335*	.882**	.690**	--							
5.	SAEBRS Emotional	-.130	.841**	.630**	.628**	--						
6.	SAEBRS Impulsive	-.313	.692**	.837**	.560**	.448**	--					
7.	Progressive Ratio Arbitrary	-.035	-.050	-.075	.079	-.024	-.011	--				
8.	Progressive Ratio Academic	.179	-.214	-.246	-.031	-.232	-.269	.542**	--			
9.	DD Real Prizes	.430*	-.136	-.093	-.285	.058	-.225	.008	-.142	--		
10.	DD Hypothetical Money	.312	-.204	-.005	-.464*	-.032	-.034	-.098	-.441*	.427*	--	
11.	DD Hypothetical Prizes	.302	-.296	-.108	-.519**	-.115	-.109	.058	-.408	.523**	.884**	--

Note. SAEBRS = Social, Academic, and Emotional Behavioral Risk Screener (Kilgus et al., 2014); DD = Delay Discounting.
* indicates a correlation significant at $\alpha = .05$
** indicates a correlation significant at $\alpha = .01$

Results of the D study are shown in Figure 4. Because task type explained a higher percentage of the variance in scores than sessions, adding tasks to the protocol would improve score dependability more than adding sessions. The line graph in Figure 4 shows the effects of increasing the number of tasks in a session on absolute *g* when scores are averaged across one, two, three, and five sessions. The dashed horizontal line represents the .7 dependability criterion. An absolute *g* greater than .7 can be achieved in several arrangements, including two sessions with five tasks, three sessions with four tasks, and five sessions with three tasks.

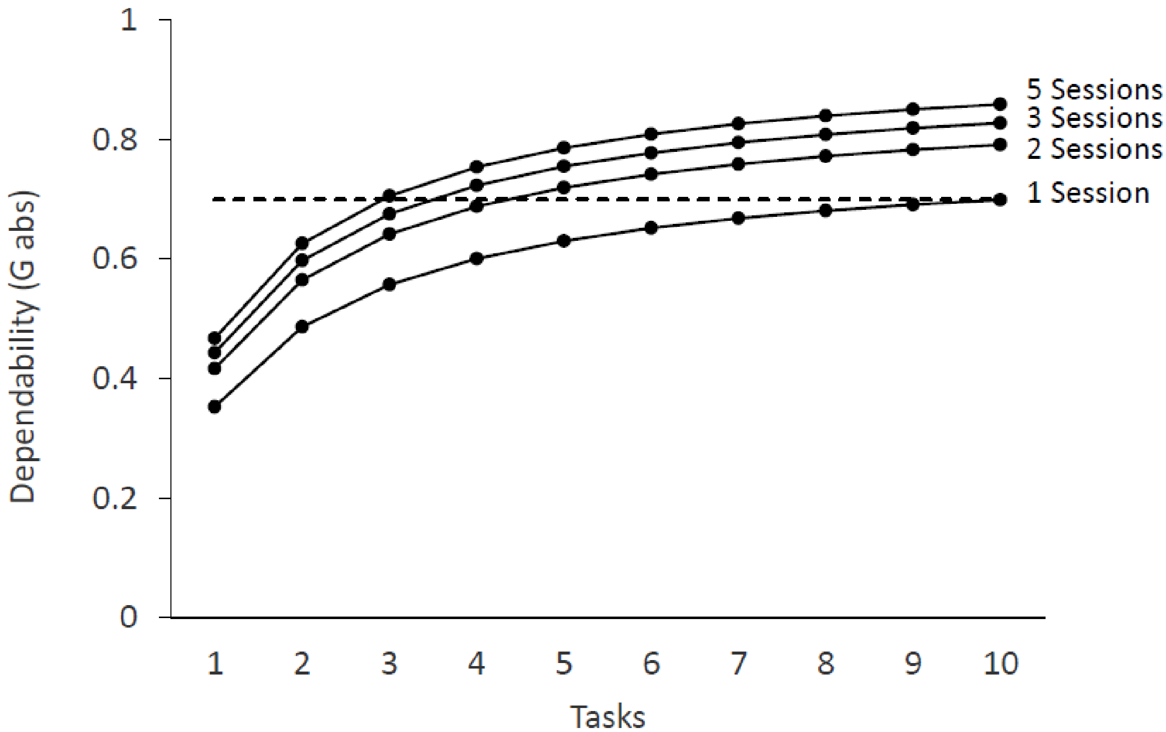


Figure 4. A demonstration of the change in absolute *g* when tasks are added to the progressive ratio assessment protocol

Research Question 2: Distribution and Correlations for Delay Discounting Scores

The distributions of *k* values obtained in all three delay discounting assessments are shown in Figure 5. The minimum and maximum possible scores are shown as black vertical lines on each distribution. Scores from students in the problem behavior group are represented by open circles, and those from the comparison group are represented by closed circles. In each assessment, several students selected the maximum delay in all five trials, resulting in a cluster of scores at the lower bound of the assessment. This may be evidence of a ceiling effect, which could conceal individual differences that would be observed if longer delays were presented. Scores were spread across the range of possible *k* values and 26 of the 30 possible *k* values were obtained by at least one student in at least one assessment. The box plots in Figure 6 show that log(*k)* values from each assessment had similar and overlapping inter-quartile ranges. The spread of scores was tightest in the real prizes assessment (*SD* = .81) and the most skew was evident in the hypothetical prizes assessment. As shown in Figure 3, all three delay discounting assessments produced similar means.

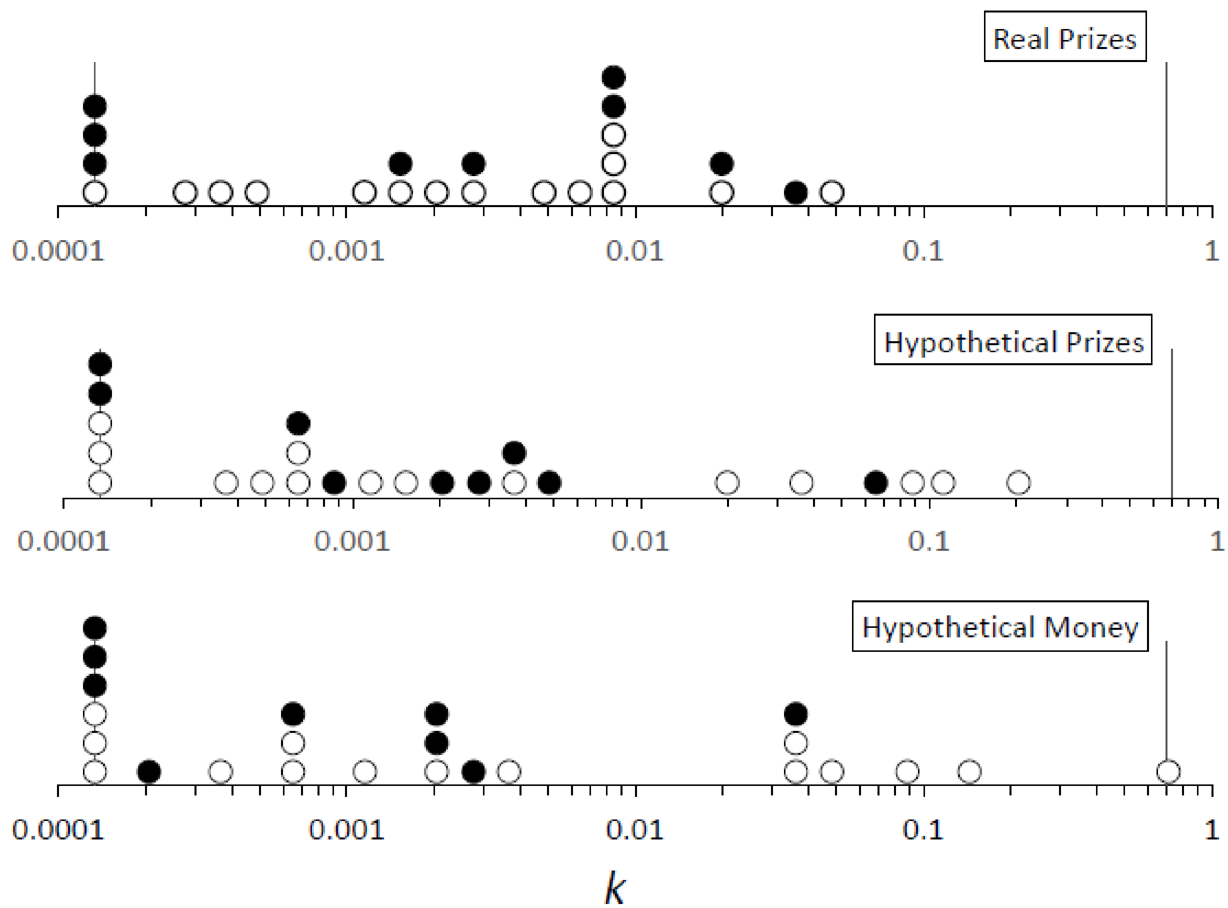


Figure 5. The distribution of *k* values obtained from the ADT-5 delay discounting assessment using real prizes (top), hypothetical prizes (middle) and hypothetical money (bottom) as rewards. Vertical lines represent the bounds of possible *k* values. Students with teacher-reported problem behavior are represented by open circles, and matched peers are represented by closed circles.

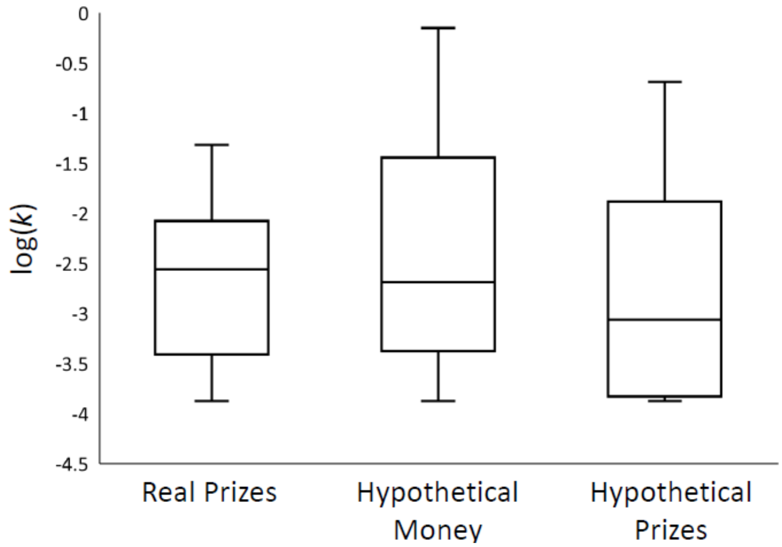


Figure 6. Box plots of log(*k*) values obtained from the ADT-5 delay discounting assessment using real prizes (left), hypothetical money (middle), and hypothetical prizes (right) as rewards.

Correspondences between *k* values obtained through different assessments are shown in Figure 7. The ranks of discounting rates estimated in the two hypothetical delay discounting assessments were strongly correlated, $\rho(22) = .85$, $p < .01$. Within student, when discounting rates for the two commodities were not equal, higher rates of discounting for money than for prizes was the most common direction of difference. The correlation between ranks of discounting rates for real and hypothetical prizes was also significant, $\rho(22) = .64$, $p < .01$. Cases of non-correspondence were most often due to students choosing more impulsively when prizes were real than when they were hypothetical. The TOST comparing real and hypothetical prizes did not return two significant *t* values ($t_L(23) = 1.803$, $p < .05$; $t_U(23) = .383$, $p > .05$), thus, I cannot reject the null hypothesis that the means of the two tests were meaningfully different. Discounting of real and hypothetical prizes are not equivalent. The TOST comparing hypothetical monetary and non-monetary rewards also did not return two significant *t* values ($t_L(23) = -1.617$, $p > .05$; $t_U(23) = -4.091$, $p < .05$), thus, I cannot reject the null hypothesis that the means of two tests were meaningfully different. Discounting on monetary and non-monetary rewards are not equivalent.

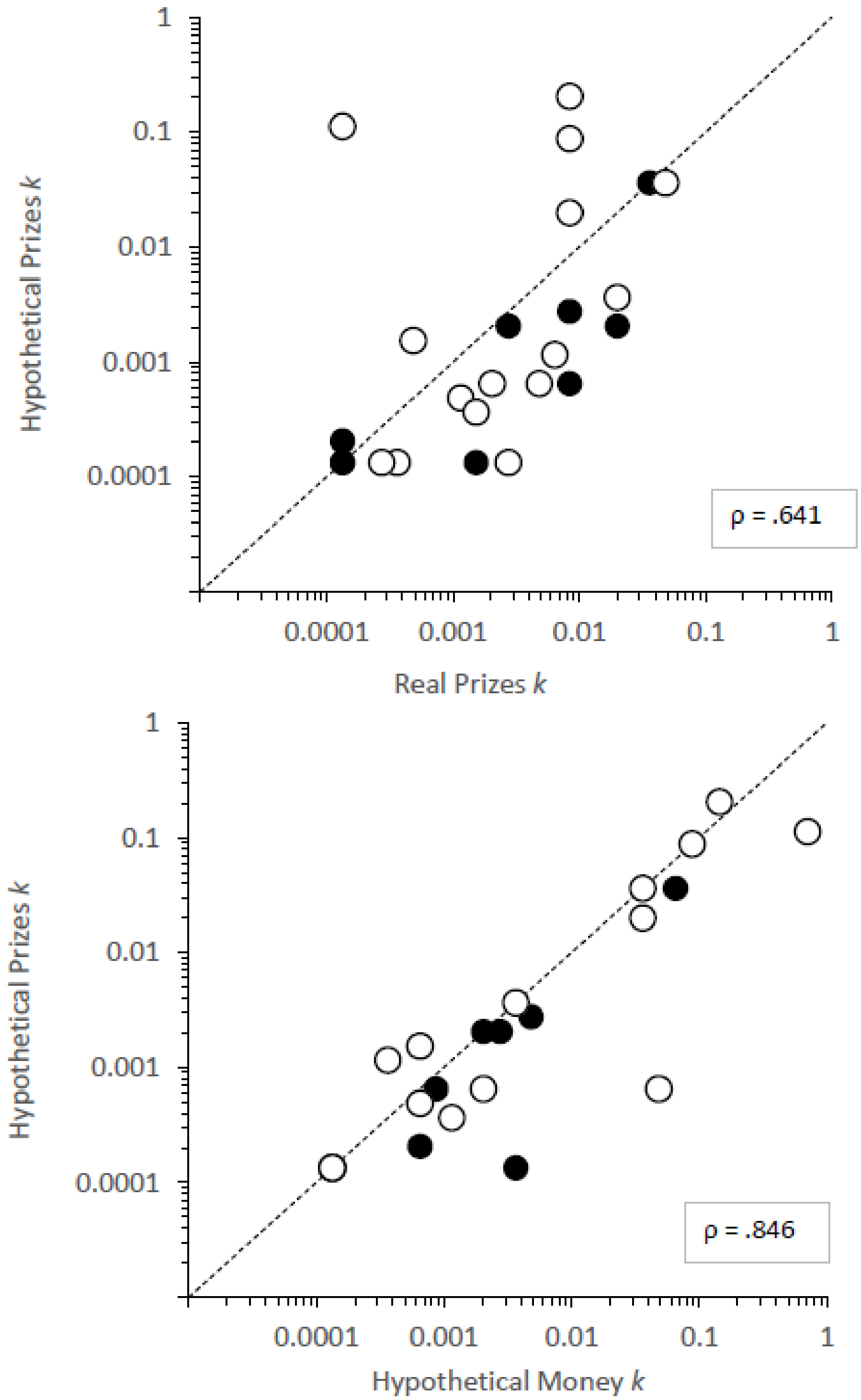


Figure 7. Associations between *k* values from the ADT-5 delay discounting assessments with hypothetical and real prizes (top) and hypothetical prizes and hypothetical money (bottom). Students with teacher-reported problem behavior are represented by open circles, and matched peers are represented by closed circles. The dashed diagonal lines show the points at which results from the two assessments are equal.

Research Question 3: Associations Between Breakpoint and Problem Behavior

Correlations between study variables are shown in Table 8. The correlation between total SAEBRS score and breakpoint determined with the arbitrary task was near zero, $r(33) = -.05$, $p > .05$. The correlation between total SAEBRS score and breakpoint determined with the academic task was larger, but not significant, $r(31) = -.214$, $p > .05$. Thus, there is not evidence to conclude there is an association between breakpoint and problem behavior. The correlations between SAEBRS total scores and the two breakpoints were not significantly different ($z = 0.95$, $p > .05$), suggesting no evidence that the relationship between breakpoint and problem behavior depends on the task used in the assessment.

Total reinforcers earned across participants at each schedule are shown in Figure 8 for the whole sample and for subgroups. When reinforcers earned were totaled across tasks, the P_{MAX} was similar for the problem behavior group ($P_{MAX} = 27$) and the non-problem behavior group ($P_{MAX} = 25$). That is, the demand curve suggested that responding would remain steady despite schedule increases until the schedule reached at least 27 responses for the problem behavior group, and at least 25 for the non-problem behavior group. Above those points, responding would decrease with every increase of response requirement. For the group of students with teacher-reported problem behavior, the estimated P_{MAX} for the arbitrary task ($P_{MAX} = 51$) was substantially higher than for the academic task ($P_{MAX} = 29$). A similar pattern was observed in the group without teacher-reported problem behavior (arbitrary $P_{MAX} = 31$; academic $P_{MAX} = 29$), though the difference was smaller. This pattern suggests that responding to the academic task was sensitive to changes in schedule requirement at lower prices relative to the arbitrary task for children with and without problem behavior, though this effect was larger for the students with problem behavior.

Research Question 4: Associations Between Delay Discounting and Problem Behavior

As shown in Table 9, I observed weak and non-significant correlations between total SAEBRS scores and results from the real delay discounting assessment, $r(22) = -.136$, $p > .05$, the hypothetical money

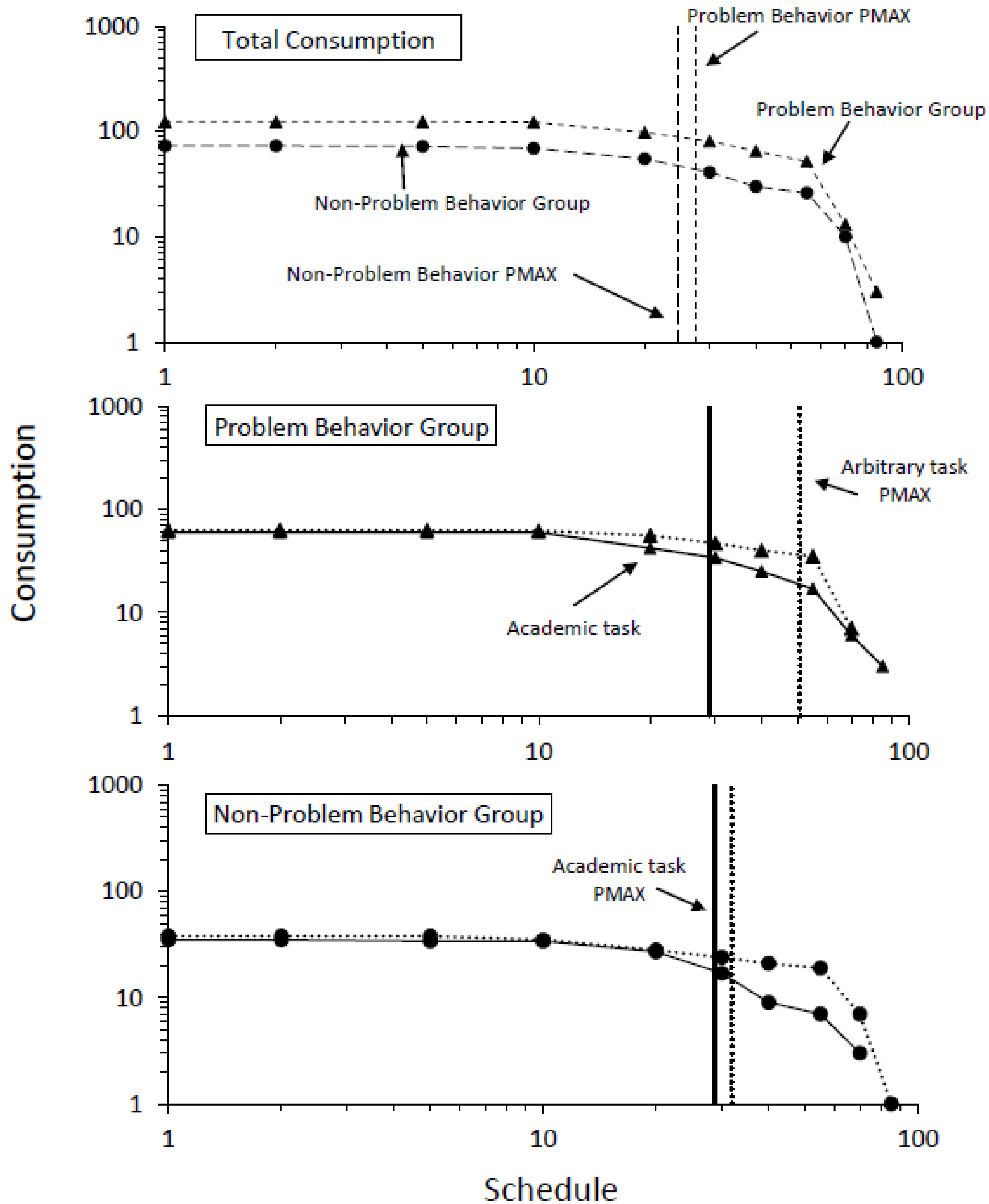


Figure 8. Graphs showing consumption as a function of schedule in the progressive ratio assessments when consumption is totaled across tasks (top) and by task for the group with teacher reported problem behavior (middle) and the group without teacher reported problem behavior (bottom). Vertical lines represent P_{MAX}.

delay discounting assessment, $r(22) = -.204$, $p > .05$, and the hypothetical prizes delay discounting assessment, $r(22) = -.296$, $p > .05$. While associations were in the expected direction (i.e., as SAEBRS scores *decreased*, indicating more behavioral risk, rates of discounting *increased*), I cannot reject the null hypothesis that the association between SAEBRS scores and discounting rate is zero. I identified no significant differences when I compared the correlations between the SAEBRS and the two hypothetical assessments ($z = -.908$, $p > .05$), or the SAEBRS and the real and hypothetical prizes assessments ($z = .777$, $p > .05$). I do not have evidence to claim that the relationship between discounting rate and problem behavior depends on the reward used to assess discounting.

Research Question 5: Correlations Between Breakpoints and Delay Discounting

The correlation between the composite breakpoint and composite discounting rate was weak and not significant, $r(22) = -.169$, $p > .05$. I cannot reject the null hypothesis that the relationship between breakpoint and delay discounting is zero.

CHAPTER 4

DISCUSSION

In schools, when behavior interventions are tailored or adapted to individual students, this is usually accomplished by altering the quality of the reinforcer used in the intervention. Research suggests that response effort, reinforcement rate and delay to reinforcement could have as much (or more) influence on responding as reinforcer quality (Neef & Lutz, 2001; Neef, et al., 1994; Romani, et al., 2017). Understanding the extent to which students are sensitive to these other dimensions of reinforcement could improve intervention design and increase flexibility for school staff. However, little is known about how to assess individual differences in sensitivity to these dimensions, and whether such differences explain classroom behavior. The purpose of this study was to pilot two behavioral assessments for young children: a progressive ratio assessment of the effects of response effort and reinforcement rate on responding and a delay discounting assessment of the effects of delay to reinforcement on responding.

Progressive Ratio

For the progressive ratio assessment, the results of a G and D study suggested that task type explained a larger proportion of variance in breakpoint scores relative to session, and that an adequately stable ranking of individuals can be achieved from progressive ratio assessments using five tasks in each of two sessions. Other arrangements that would produce dependable results were four tasks in each of three sessions, and three tasks in each of five sessions. In this study, completing one of the tasks took an estimated average of 7.4 min (range, 0.3 min-19.2 min). Based on these averages, a 2-session, 5-task protocol would take an average of 37 min per session or 74 min total, to complete. A 4-task, 3-session protocol would take 89 min and a 3-task, 5-session protocol would take 111 min to complete.

Recommendations for a progressive ratio assessment protocol based on these results vary based on context. In the context of a research study, it might be more difficult to use more tasks in fewer sessions than to repeat more sessions with fewer tasks. For each new task, researchers must design responses, train implementers, and manage data collection, which could be more effortful than adding sessions. In schools, however, if it is important to minimize time away from instruction, five tasks repeated twice may be the more appropriate arrangement, as it is the protocol with the shortest overall duration. At an estimated average of 74 total min, this protocol is comparable to other behavioral and academic tests used in schools. For example, the Functional Assessment Interview (O'Neill et al., 1997) takes an estimated 45-90 min to complete, and MAP Growth assessments take the average student 40-44 min per each of three subtests in the lower grades (Northwest Evaluation Association, 2018). A standard functional analysis (see Beavers, Iwata, & Lerman [2013] for review)with a minimum of three 5-min sessions in each of four conditions (e.g., play, tangible, attention and escape) would take 60 min, with some literature reviews suggesting much longer durations to achieve differentiated results (e.g., a mean of 4 h in Lloyd, Weaver, & Staubitz, 2016)

With respect to whether progressive ratio results could explain classroom behavior, I did not find evidence that children with problem behavior were more sensitive to changes in reinforcement schedule than children without problem behavior. Not only were correlations between SAEBRS total score and progressive ratio breakpoint not significant, they were in an unexpected direction: as SAEBRS scores decreased (indicating higher risk), breakpoints increased, especially in the academic task. Nor did I find evidence that the association between problem behavior and breakpoint was significantly different when arbitrary versus academic tasks were used in the assessment. This suggests that progressive ratio assessments are not suited to identify students who are likely to engage in problem behavior. However, this finding does not rule out their validity for making other kinds of inferences (e.g., to describe mismatches between a student's environment and his or her tolerance for thin schedules of reinforcement, or to select or modify behavior interventions for individual children).

Delay Discounting

The second assessment was a delay discounting assessment to measure the rate at which future rewards lose value for young children. I found some evidence of a ceiling effect in the ADT-5 delay discounting assessment, though many scores were spread across the range of possible values. This suggests that some sensitivity in distinguishing between children with low rates of discounting is lost in the ADT-5 because it uses short delays. However, we do not know whether those distinctions would be meaningful or useful. It could be the case that the variability obscured by the ceiling effect observed here is important for the measure's ability to predict behavior outside of the assessment context. It is just as likely that the meaningful variability is at the center of the scale, not at one extreme.

In assessments using hypothetical rewards, there's little downside to using a longer maximum delay to increase the range of individual difference that could be detected. But when rewards are real, longer delays may complicate procedures, especially if delays are long enough to exceed the school day. Given the assessments using real and hypothetical rewards were not equivalent, it may be important to determine whether one or the other produces more valid and reliable information, before deciding whether to modify the assessment to include longer delays. Future work on the validity of the ADT-5 with real and hypothetical rewards, and with the inclusion of longer delays, could help identify the assessment with the most meaningful sensitivity.

Though the different ADT-5 formats were not equivalent, I did find substantial correspondence between discounting rates across reward type and status. This is consistent with the theory that all versions of the discounting assessment are measuring the same construct, but that the reward type and/or status has an effect on discounting rates. Within student, when discounting for the two commodities (money and prizes) were not equal ($n = 15$), higher rates of discounting for money than for prizes was the most common direction of difference ($n=12$). In adults, money is often discounted less than other kinds of rewards (e.g., Baker, Johnson, & Bickel, 2003; Koffarnus & Bickel, 2014; Weatherly, Terrell, & Derenne, 2010). There is no evidence of a similar effect in the present study. This suggests that money discounting functions differently in young children than it does in adults. This could be because children have less experience with money and it does not function effectively as a conditioned reinforcer.

When students gave different responses on the ADT-5 when prizes were real than when they were hypothetical (n = 21), it was most often because students chose more impulsively when real prizes were at stake (n = 17). The correlation I found between real and hypothetical discounting was lower than those found in adults (e.g., Johnson & Bickel, 2002; Madden et al., 2003). This may be because young children are less able to imagine hypothetical situations than adults. The correlation between real and hypothetical assessments found in this study was also lower than what Miller (2019) identified for children. This could be because Miller assessed discounting of money, and the difference between the real item and its verbal description was more salient for prizes than for money. It also could be because the mean age in Miller's study was higher than our sample (i.e., 9 years vs 7 years 9 months), and older children may be better equipped to imagine hypotheticals than younger ones. I also excluded children from this study if they expressed a preference for delayed rewards when amounts were equal at pre-test. I hypothesized that such responding was evidence of social desirability bias, and that those students might choose delayed rewards consistently, not because of the subjective value of the reward, but to please the assessor. If such students had been included and did select delayed rewards across assessments, that would have increased the correlations between assessment types. Exclusion of such children may be one reason that the correlation in the present study was lower than the one Miller found.

I did not find evidence of a significant relationship between problem behavior and discounting. However, the non-significant correlations between total SAEBRS scores and discounting rates were in the expected direction: as SAEBRS scores decreased (indicating more risk), discounting rate increased. I did find significant correlations between hypothetical discounting and the SAEBRS Academic subscale (shown in Table 8). The Academic subscale reflects teacher ratings of: a) interest in academic topics; b) preparedness for instruction; c) production of acceptable work; d) difficulty working independently; e) distractedness and e) academic engagement. These items might help to define the construct measured in the ADT-5 for young children. Overall, I found no evidence that relationship between problem behavior and discounting rate depended on the reward used in the session.

Limitations

The results of this study should be interpreted in light of several important limitations. First, there were significant and potentially meaningful differences between the students recruited because they had problem behavior and the students recruited to represent peers without problem behavior. This is because teachers were not able to identify a matched peer for every student with problem behavior who agreed to participate. Older students and students with low SES were less likely to be matched. Therefore, students in the comparison group without problem behavior were younger, on average, and less likely to qualify for free or reduced lunch than the students with problem behavior. Thus, age was significantly correlated with SAEBRS total score, $r(33) = -.344$, $p < .05$, and students eligible for FRL had significantly lower mean SAEBRS total scores than students not eligible for FRL, $t(33) = 2.5$, $p < .05$. Any associations between assessment variables and SAEBRS scores could be due to differences in age or SES.

In this sample, however, there were no significant correlations between age and academic task breakpoint, $r(31) = .075$, $p > .05$, arbitrary task breakpoint, $r(33) = -.087$, $p > .05$, or composite breakpoint, $r(33) = .11$, $p > .05$. Similarly, mean academic and arbitrary breakpoints were not significantly or meaningfully different for students who did and did not qualify for FRL. Therefore, differences in age and FRL eligibility are unlikely to confound conclusions about associations between breakpoints and SAEBRS scores. There was a significant correlation between age and the log(*k)* from the real delay discounting assessment, $r(22) = .43$, $p < .05$. Additionally, students eligible for FRL had significantly higher discounting rates in the real, $t(22) = -3.02$, $p < .01$, and hypothetical prizes $t(22) = -2.33$, $p < .05$ assessments. Because age or FRL eligibility could be confounds in my analysis of the relationship between discounting rates and problem behavior, I re-examined these correlations controlling for age and FRL eligibility. Correlations between total SAEBRS scores and delay discounting from real, $r(20) = .024$, $p > .05$, hypothetical money,

$r(20) = -.108$, $p > .05$, and hypothetical prizes, $r(20) = -.206$, $p > .05$ with these controls were similar to those found in the whole sample: weak, negative and non-significant.

A second limitation is that assessment procedures may have affected the results of the progressive ratio assessments. To minimize the time students were pulled from class, I set a limit of 10 minutes per task, and interrupted students at the next reinforcement delivery if they were still working after that limit. This could have introduced a ceiling effect, and limited my ability to identify individual differences among students with high breakpoints. However, every student reached a point where they requested to stop working in at least one progressive ratio assessment. That is, no student's composite breakpoint score was entirely due to the time limitation.

Finally, in the arbitrary task, it was necessary at some point to empty the cubes that the student had already placed in the box, so that there would always be enough cubes to complete the next response requirement. I programmed this after the FR-55 schedule, following a cumulative 163 responses. Anecdotally, several students expressed a desire to put all the cubes in the box during the assessment, and several expressed mild disappointment when the box was emptied. It is possible that this influenced some students' breakpoints. If they were motivated to keep responding as they saw the number of cubes left to put in the box dwindle, or less motivated when the box was emptied, they might request to end the assessment after the FR-55 schedule. In fact, 55 was the most frequent breakpoint obtained in the arbitrary task.

Similarly, in the academic task, students were more likely to stop responding at the end of a worksheet than would be predicted by chance alone. About half of the academic progressive ratio sessions ended after a student completed a full worksheet. There was no schedule that aligned with the end of a worksheet, and the tendency to complete worksheets does not seem to have made any one breakpoint more likely than the others. However, in all but two sessions, students completed at least one worksheet. The first worksheet was completed after the FR-10 reinforcer, but before the FR-20 reinforcer. Thus, demand in the academic task remaining inelastic until FR-20 might be related in part to a rule-governed tendency to complete whole worksheets.

Though I had an inclusion criterion for the delay discounting assessment that I used to check for and exclude children whose responses were likely biased by social desirability, I had no such criteria for inclusion in the progressive ratio assessment. The desire to finish a worksheet or to complete the arbitrary task by getting all the blocks in the box may be related to a desire to please the assessor. One child asked during the academic task "Won't you be proud of me if I finish all the math?" Though implementers were trained to minimally interact with students and maintain a neutral affect during the assessment, students in this sample likely had a substantial learning history of receiving praise for completing tasks and working diligently.

Since worksheet length and programmed box emptying were consistent across sessions, these factors might have increased the consistency of breakpoints across sessions and within task for some students. If many students were influenced in this way, that might decrease the overall variability in breakpoint within task and across students. Both of these could have inflated our estimated *g* coefficient. However, both effects would have increased the variability between tasks, which would have the opposite effect on *g*.

Future Directions

In this study, I looked at the ways that assessment parameters might affect student responding. Though I probed associations between both breakpoint and *k* and behavioral risk as measured by the SAEBRS, this study did not substantially address questions of validity. Future studies are needed to evaluate the concurrent and predictive validity of these assessments. Before these assessments can be recommended for use by school staff to select or individualize behavior interventions, research is needed on the extent to which ATD-5 or progressive ratio results predict responses to interventions and explain classroom behavior.

Further, results of the present study raise questions about the traits reflected in these assessments. Though in the delay discounting literature, a preference for an immediate reward is often interpreted as an impulsive choice (e.g., Madden & Johnson, 2010), correlations between all delay discounting measures and the impulsiveness item on the SAEBRS were weak and insignificant. This suggests that '*impulsiveness'* as defined by a preference for immediate reinforcers and the construct teachers understand as '*impulsiveness'* could be meaningfully different. As the Academic subscale of the SAEBRS significantly correlated with the discounting measures, future work examining the relationship of discounting rate and '*distractedness*' or '*difficulty working independently*' could be fruitful.

Especially for progressive ratio assessments, developing computer- or tablet-aided assessment software could be beneficial. If task responses were programmed on tablets, data collection could be automated, and influences from assessment details (like the motivation to empty the bin or the desire to complete worksheets during sessions) could be eliminated. Progressive ratio assessments conducted as described in this study require sustained one-on-one attention, which can be a strain on resources in research and in practice. Larger scale studies of progressive ratio validity might be more easily accomplished using tablets or computers, and these assessments might be more acceptable and feasible in schools. Pilot research would be needed to develop and optimize such electronic assessments, and to determine their social validity.

Finally, in this study, we constructed demand curves for two different tasks, and compared the P_{MAX} from each. Like progressive ratio assessments, comparisons of P_{MAX} are most often done to compare the effects of different reinforcers on responding. However, this may be a way to quantify the effort needed to complete different responses. Conceptually, it seems sound to assume that performance of an effortful response would be more sensitive to smaller changes in schedule than a less effortful response. Comparisons of P_{MAX} could be useful for teachers, clinicians, or researchers who need to identify the relative effort involved in a response. The extent to which this kind of comparison is useful in research or practice should be evaluated in future studies.

In conclusion, both delay discounting and progressive ratio assessments have promise for explaining the way that dimensions of reinforcement other than quality impact student behavior. Findings in this study suggest these assessments can produce reliable results, and that they function similarly in children with and without problem behavior. Future research evaluating the concurrent and predictive validity of these assessments could pave the way towards helping teachers and other school staff select or individualize behavior interventions.

REFERENCES

Ainslie, G. W. (1974). Impulse control in pigeons. *Journal of the Experimental Analysis of Behavior, 21*, 485-489.

Audrain-McGovern, J., Rodriguez, D., Epstein, L. H., Cuevas, J., Rodgers, K., & Wileyto, E. P. (2009). Does delay discounting play an etiological role in smoking or is it a consequence of smoking? *Drug and Alcohol Dependence, 103*, 99-106.

Baker, F., Johnson, M. W., & Bickel, W. K. (2003). Delay discounting in current and never-before cigarette smokers: similarities and differences across commodity, sign, and magnitude. *Journal of Abnormal Psychology, 112*, 382-392.

Beavers, G. A., Iwata, B. A., & Lerman, D. C. (2013). Thirty years of research on the functional analysis of problem behavior. *Journal of Applied Behavior Analysis, 46*, 1-21.

Best, J. R., Theim, K. R., Gredysa, D. M., Stein, R. I., Welch, R. R., Saelens, B. E., ... & Wilfley, D. E. (2012). Behavioral economic predictors of overweight children's weight loss. *Journal of Consulting and Clinical Psychology, 80*, 1086-1096.

Broussard, C., & Northup, J. (1997). The use of functional analysis to develop peer interventions for disruptive classroom behavior. *School Psychology Quarterly*, *12*, 65-76. http://dx.doi.org/10.1037/h0088948

Billington, E., & DiTommaso, N. M. (2003). Demonstrations and applications of the matching law in education. *Journal of Behavioral Education, 12*, 91-104.

Chelonis, J. J., Gravelin, C. R., & Paule, M. G. (2011). Assessing motivation in children using a progressive ratio task. *Behavioural Processes, 87,* 203-209.

DeLeon, I. G., Frank, M. A., Gregory, M. K., & Allman, M. J. (2009). On the correspondence between preference assessment outcomes and progressive-ratio schedule assessments of stimulus value. *Journal of Applied Behavior Analysis, 42*, 729-33.

Dougherty, D. M., Charles, N. E., Mathias, C. W., Ryan, S. R., Olvera, R. L., Liang, Y., & Acheson, A. (2014). Delay discounting differentiates pre-adolescents at high and low risk for substance use disorders based on family history. *Drug and Alcohol Dependence*, 143, 105-111.

Duckworth, A. L., & Seligman, M. E. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science, 16*, 939-944.

DuPaul, G. J., & Stoner, G. (2014). *ADHD in the schools: Assessment and intervention strategies*. Guilford Publications. New York

Francisco, M. T., Borrero, J. C., & Sy, J. R. (2008). Evaluation of absolute and relative reinforcer value using progressive-ratio schedules. *Journal of Applied Behavior Analysis*, *41*, 189-202.

Gilroy, S. P., Kaplan, B. A., Reed, D. D., Koffarnus, M. N. & Hantula, D. A. (2018). The Demand Curve Analyzer: Behavioral economic software for applied researchers. *Journal of the Experimental Analysis of Behavior*. doi: 10.1002/jeab.479

Glover, A. C., Roane, H. S., Kadey, H. J., & Grow, L. L. (2008). Preference for reinforcers under progressive-and fixed-ratio schedules: A comparison of single and concurrent arrangements. *Journal of Applied Behavior Analysis*, *41*, 163-176.

Hagopian, L. P., Long, E. S., & Rush, K. S. (2004). Preference assessment procedures for individuals with developmental disabilities. *Behavior Modification, 28*, 668-677.

Hall, R. V., Lund, D., & Jackson, D. (1968). Effects of teacher attention on study behavior. *Journal of Applied Behavior Analysis, 1*, 1-12.

Harris, F. R., Wolf, M. M., & Baer, D. M. (1964). Effects of adult social reinforcement on child behavior. *Young Children, 20*, 8-17.

Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior, 4,* 267–272.

Hershenberg, R., Satterthwaite, T. D., Daldal, A., Katchmar, N., Moore, T. M., Kable, J. W., & Wolf, D. H. (2016). Diminished effort on a progressive ratio task in both unipolar and bipolar depression. *Journal of Affective Disorders, 196*, 97-100.

Horner, R. H., Sugai, G., & Anderson, C. M. (2010). Examining the evidence base for school-wide positive behavior support. *Focus on Exceptional Children, 42,* 1-14.

Johnson, M. W., & Bickel, W. K. (2002). Within-subject comparison of real and hypothetical money rewards in delay discounting. *Journal of the Experimental Analysis of Behavior, 77*, 129-146.

Kilgus, S. P., Chafouleas, S. M., Riley-Tillman, T. C., & von der Embse, N. P. (2014). Social, Academic, and Emotional Behavior Risk Screener (SAEBRS). Minneapolis, MN: Theodore J. Christ & Colleagues.

Kilgus, S. P., Eklund, K., Nathaniel, P., Taylor, C. N., & Sims, W. A. (2016). Psychometric defensibility of the Social, Academic, and Emotional Behavior Risk Screener (SAEBRS) Teacher Rating Scale and multiple gating procedure within elementary and middle school samples. *Journal of School Psychology, 58*, 21-39.

Kirby, K. N., & Marakovic, N. N. (1996). Delay-discounting probabilistic rewards: Rates decrease as amounts increase. *Psychonomic Bulletin & Review, 3*, 100–104. doi:10.3758/BF03210748.

Kirby, K. N., Petry, N. M., & Bickel, W. K. (1999). Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. *Journal of Experimental Psychology: General, 128*, 78–87. doi:10.1037/0096-3445.128.1.78.

Kodak, T., Lerman, D. C., & Call, N. (2007). Evaluating the influence of postsession reinforcement on choice of reinforcers. *Journal of Applied Behavior Analysis, 40*, 515-527.

Koffarnus, M. N., & Bickel, W. K. (2014). A 5-trial adjusting delay discounting task: Accurate discount rates in less than one minute. *Experimental and Clinical Psychopharmacology, 22*, 222-228.

Kollins, S. H. (2003). Delay discounting is associated with substance use in college students. *Addictive Behaviors, 28*, 1167-1173.

Lakens, D. (2017). Equivalence tests: a practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science, 8*, 355-362.

Lee, I. A., & Preacher, K. J. (2013, September). Calculation for the test of the difference between two dependent correlations with one variable in common [Computer software]. Available from http://quantpsy.org/corrtest/corrtest2.htm

Mace, F. C., McCurdy, B., & Quigley, E. A. (1990). A collateral effect of reward predicted by matching theory. *Journal of Applied Behavior Analysis, 23*, 197-205.

Lloyd, B. P., Weaver, E. S., & Staubitz, J. L. (2016). A review of functional analysis methods conducted in public school classroom settings. *Journal of Behavioral Education, 25*, 324-356.

Madden, G. J., Begotka, A. M., Raiff, B. R., & Kastern, L. L. (2003). Delay discounting of real and hypothetical rewards. *Experimental and Clinical Psychopharmacology, 11*, 139-145.

Madden, G. J. & Johnson, P. S. (2010). A delay-discounting primer. In G. J. Madden & W. K. Bickel (Eds.), *Impulsivity: The behavioral and neurological science of discounting* (pp. 11-37). Washington, DC: American Psychological Association.

Maggin, D. M., Chafouleas, S. M., Goddard, K. M., & Johnson, A. H. (2011). A systematic evaluation of token economies as a classroom management tool for students with challenging behavior. *Journal of School Psychology, 49*, 529-554.

Maggin, D. M., Johnson, A. H., Chafouleas, S. M., Ruberto, L. M., & Berggren, M. (2012). A systematic evidence review of school-based group contingency interventions for students with challenging behavior. *Journal of School Psychology, 50*, 625-654.

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. In: M. L. Commons, J. E. Mazur, J. A. Nevin, & H. Rachlin (Eds.). *Quantitative analysis of behavior* (Vol. 5, pp. 55-73) Hillsdale, NJ: Erlbaum.

McLeod, D. R., & Griffiths, R. R. (1983). Human progressive-ratio performance: Maintenance by pentobarbital. *Psychopharmacology, 79*, 4-9.

Miller, J. R. (2019). Comparing rapid assessments of delay discounting with real and hypothetical rewards in children. *Journal of the Experimental Analysis of Behavior, 111*, 48-58.

Miras, A. D., Jackson, R. N., Jackson, S. N., Goldstone, A. P., Olbers, T., Hackenberg, T., ... & le Roux, C. W. (2012). Gastric bypass surgery for obesity decreases the reward value of a sweet-fat stimulus as assessed in a progressive ratio task. *The American Journal of Clinical Nutrition, 96*, 467-473.

Neef, N. A., & Lutz, M. N. (2001). A brief computer-based assessment of reinforcer dimensions affecting choice. *Journal of Applied Behavior Analysis, 34*, 57-60.

Neef, N. A., Shade, D., & Miller, M. S. (1994). Assessing influential dimensions of reinforcers on choice in students with serious emotional disturbance. *Journal of Applied Behavior Analysis, 27*, 575-583.

Northwest Evaluation Association (2018). Average MAP growth test durations. Retrieved from https://www.nwea.org/content/uploads/2018/08/Average-MAP-Growth-Test-Durations.pdf

Odum, A. L. (2011a). Delay discounting: I'm a k, you're a k. *Journal of the Experimental Analysis of Behavior, 96*, 427-439.

Odum A. L. (2011b). Delay discounting: Trait variable? *Behavioural Processes, 87*, 1-9.

Odom, S. L., Collet-Klingenberg, L., Rogers, S. J., & Hatton, D. D. (2010). Evidence-based practices in interventions for children and youth with autism spectrum disorders. *Preventing School Failure: Alternative Education for Children and Youth, 54*, 275-282.

O'Leary, K. D., Becker, W. C., Evans, M. B., & Saudargas, R. A. (1969). A token reinforcement program in a public school: A replication and systematic analysis. *Journal of Applied Behavior Analysis, 2*, 3-13.

O'Neill, R. E., Horner, R. H., Albin, R. W., Sprague, J. R., Storey, K., & Newton, J. S. (1997). *Functional assessment and program development for problem behavior: A practical handbook*. Pacific Grove, CA: Brooks/Cole.

Paclawskyj, T. R., & Vollmer, T. R. (1995). Reinforcer assessment for children with developmental disabilities and visual impairments. *Journal of Applied Behavior Analysis, 28*, 219-224.

Paramore, N. W., & Higbee, T. S. (2005). An evaluation of a brief multiple-stimulus preference assessment with adolescents with emotional-behavioral disorders in an educational setting. *Journal of Applied Behavior Analysis, 38*, 399-403.

Reed, D. D., Luiselli, J. K., Magnuson, J. D., Fillers, S., Vieira, S., & Rue, H. C. (2009). A comparison between traditional economical and demand curve analyses of relative reinforcer efficacy in the validation of preference assessment predictions. *Developmental Neurorehabilitation, 12*, 164-169.

Reed, D. D., & Martens, B. K. (2011). Temporal discounting predicts student responsiveness to exchange delays in a classroom token system. *Journal of Applied Behavior Analysis, 44*, 1-18.

Reynolds, B. (2006). A review of delay-discounting research with humans: relations to drug use and gambling. *Behavioural Pharmacology, 17*, 651-667.

Roane, H. S. (2008). On the applied use of progressive-ratio schedules of reinforcement. *Journal of Applied Behavior Analysis, 41*, 155-161.

Roane, H. S., Lerman, D. C., & Vorndran, C. M. (2001). Assessing reinforcers under progressive schedule requirements. *Journal of Applied Behavior Analysis, 34*, 145-167.

Romani, P. W., Alcorn, A. S., Miller, J. R., & Clark, G. (2017). Preference assessment for dimensions of reinforcement to inform token economies targeting problem behavior. *Journal of Behavioral Education, 26*, 221-237.

Shavelson, R. J., & Webb, N. M. (2006). Generalizability Theory. In J. L. Green, G. Camilli, & P. B. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 309-322). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Shiels, K., Hawk Jr, L. W., Reynolds, B., Mazzullo, R. J., Rhodes, J. D., Pelham Jr, W. E., ... & Gangloff, B. P. (2009). Effects of methylphenidate on discounting of delayed rewards in attention deficit/hyperactivity disorder. *Experimental and Clinical Psychopharmacology, 17*, 291-301.

Solomon, B. G., Klein, S. A., Hintze, J. M., Cressey, J. M., & Peller, S. L. (2012). A meta-analysis of school-wide positive behavior support: An exploratory study using single-case synthesis. *Psychology in the Schools, 49*, 105-121.

Stanger, C., Ryan, S. R., Fu, H., Landes, R. D., Jones, B. A., Bickel, W. K., & Budney, A. J. (2012). Delay discounting predicts adolescent substance abuse treatment outcome. *Experimental and Clinical Psychopharmacology, 20*, 205-212.

Staubitz, J. L., Lloyd, B. P., & Reed, D. D. (2018). A summary of methods for measuring delay discounting in young children. *The Psychological Record, 68*, 239-253.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87, 245-251.

Strauss, G. P., Whearty, K. M., Morra, L. F., Sullivan, S. K., Ossenfort, K. L., & Frost, K. H. (2016). Avolition in schizophrenia is associated with reduced willingness to expend effort for reward on a progressive ratio task. *Schizophrenia Research, 170*, 198-204.

Tiger, J. H., Toussaint, K. A., & Roath, C. T. (2010). An evaluation of the value of choice-making opportunities in single-operant arrangements: Simple fixed-and progressive-ratio schedules. *Journal of Applied Behavior Analysis, 43*, 519-524.

Weatherly, J. N., Terrell, H. K., & Derenne, A. (2010). Delay discounting of different commodities. *The Journal of General Psychology: Experimental, Psychological, and Comparative Psychology, 137*, 273-286.

Wolf, D. H., Satterthwaite, T. D., Kantrowitz, J. J., Katchmar, N., Vandekar, L., Elliott, M. A., & Ruparel, K. (2014). Amotivation in schizophrenia: Integrated assessment with behavioral, clinical, and imaging measures. *Schizophrenia Bulletin, 40*, 1328-1337.

Wright-Gallo, G. L., Higbee, T. S., Reagon, K. A., & Davey, B. J. (2006). Classroom-based functional analysis and intervention for students with emotional/behavioral disorders. *Education and Treatment of Children*, *29*, 421-436.

Yi, R., Mitchell, S. H., & Bickel, W. K. (2010). Delay discounting and substance abuse-dependence. In G. J. Madden & W. K. Bickel (Eds.), *Impulsivity: The behavioral and neurological science of discounting* (pp. 191-211). Washington, DC: American Psychological Association.

Yoder, P. J., Lloyd, B. P., & Symons, F. J. (2018). *Observational measurement of behavior.* Baltimore, MD: Brookes.

Yoon, J. H., & Higgins, S. T. (2008). Turning k on its head: Comments on use of an ED50 in delay discounting research. *Drug and Alcohol Dependence, 95*, 169-172.

Appendix A

Example Math Worksheet For Progressive Ratio Academic Task

Math Worksheet

Math Fact Cafe™
© Math Fact Cafe

1 + 3	0 + 6	3 + 5	0 + 5	6 + 2
3 + 6	8 + 0	2 + 6	7 + 0	4 + 2
5 + 4	6 + 3	0 + 8	4 + 0	0 + 9
4 + 2	4 + 0	1 + 9	5 + 1	2 + 4
2 + 0	3 + 7	9 + 1	2 + 6	1 + 4
6 + 0	4 + 2	7 + 1	3 + 6	5 + 5

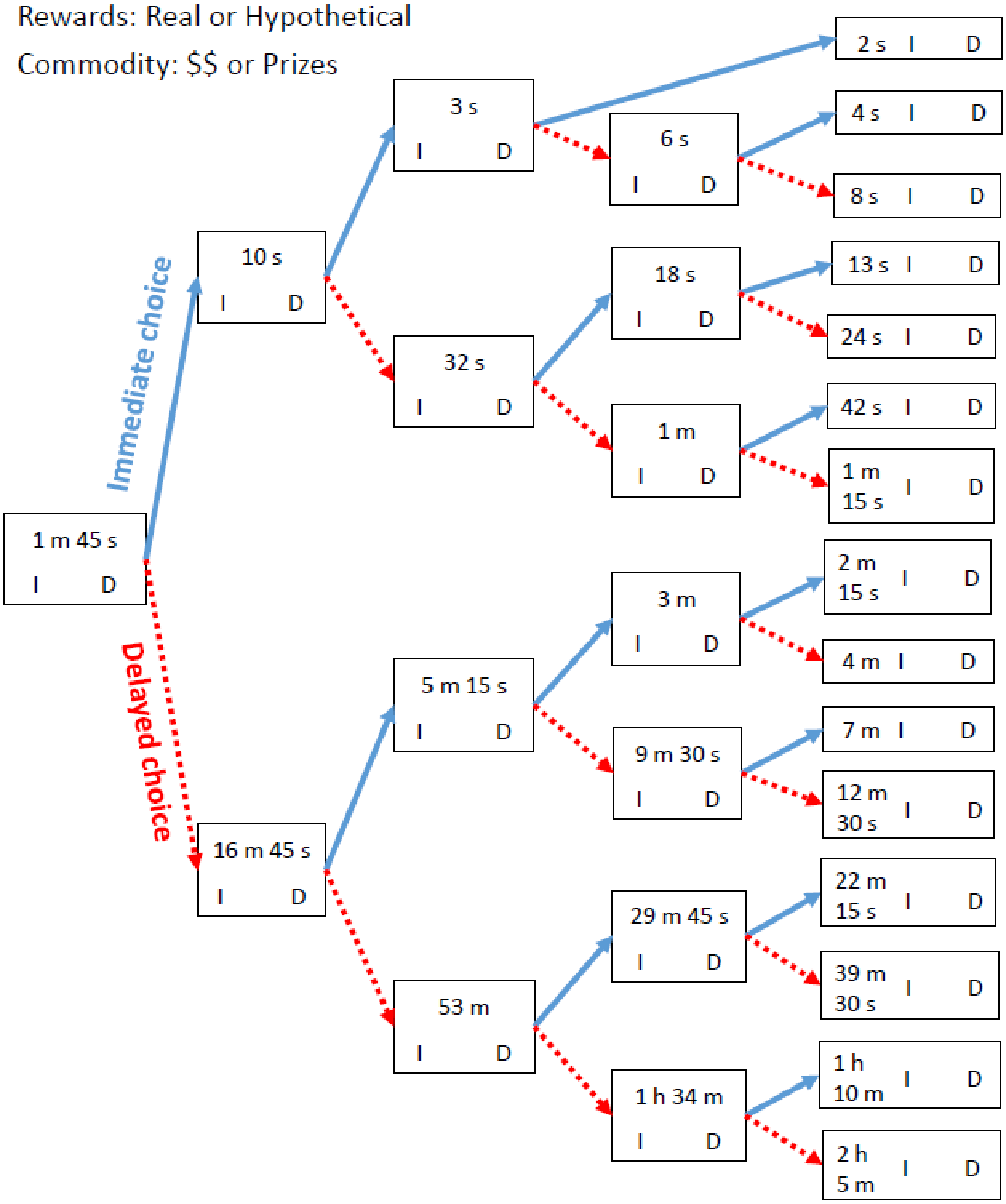
Appendix B

Example Data Collection Sheet ADT-5

Participant ID:_________ RA:_____________Date:________________ Prime Reli

Rewards: Real or Hypothetical

Commodity: $$ or Prizes



Instructions: In each choice box, circle whether participant chooses the immediate reward (I) or the delayed reward (D). Follow the arrows to determine the next choice.

Appendix C

ADT-5 Script

Real Non-Monetary Commodity:
Today, I have a treasure box with me [show box, let's briefly examine the items in it]. I'm going to ask you some questions about when you'd like to get a treat from the treasure box.

Remember, it takes 1 second to say your name, 15 seconds to say the Pledge, 1 minute to wash your hands, and 30 minutes to watch a TV show. Do you have any questions?

Do you want to pick one treat from the treasure box right now OR do you want to pick two treats from the treasure box in 1 minute and 45 seconds? [record answer]

If s chooses 'right now': OK, great, you can pick your treat now [allow s to choose one item, then present next trial]

If s chooses 'later': OK, we'll set a timer for 1 minute and 45 seconds. [show timer, set and press start, then remove timer] The timer will remind me when to get the treasure box for you so you can pick your treat.

Hypothetical Non-Monetary Commodity:
Today, I am going to ask you to imagine getting a treat from a treasure box. I'm going to ask you some questions about when you'd like to get treats and how many you'd like to get. But all the treats will be imaginary.

Remember, it takes 1 second to say your name, 15 seconds to say the Pledge, 1 minute to wash your hands, and 30 minutes to watch a TV show. Do you have any questions?

Would you want to pick one treat from the treasure box right now OR would you want to pick two treats from the treasure box in 1 minute and 45 seconds? [record answer and present next trial]

Hypothetical Monetary Commodity:
Today, I am going to ask you to imagine getting some money. I'm going to ask you some questions about when you'd like to get money and how much you'd like to get. But all the money will be imaginary.

Remember, it takes 1 second to say your name, 15 seconds to say the Pledge, 1 minute to wash your hands, and 30 minutes to watch a TV show. Do you have any questions?

Would you like to get one quarter right now OR would you like to get two quarters in 1 minute and 45 seconds? [record answer and present next trial]

Appendix D

Introductory Scripts for Progressive Ratio Assessments

Arbitrary Task Script:
"Here we have a lot of blocks. In this activity, you can put blocks in this box. You must put them in one at a time, like this [model]. When you put blocks in the box, then you can have [reinforcer]. Sometimes you will have to put a lot of blocks in to get [reinforcer], sometimes only a few. You can keep putting blocks in the jar as long as you want to keep earning [reinforcer], but you can stop any time you want. Let's try it" [example response & reward]

"Remember, you can keep putting blocks in the jar as long as you want to. If you don't want to put any more blocks in the jar, you can say "all done" or touch the STOP sign like this. Do you have any questions?"

Respond to any questions, then count out "ready, set, go", start session timer.

Academic Task Script:
"Here is some math work. In this activity, you can solve math problems. You must write your answer below the problem, like this [model]. When you do math, you can have [reinforcer]. Sometimes you will have to do a lot of problems to get [reinforcer], sometimes only a few. You can keep working on math as long as you want to keep earning [reinforcer], but you can stop any time you want. Let's try it, write an answer here" [reinforce]

"Remember, you can keep doing math as long as you want to. If you don't want to do any more math, you can say "all done" or touch the STOP sign like this. Do you have any questions?"

Respond to any questions, then count out "ready, set, go", start session timer